

---

# Universidad Complutense de Madrid

## Facultad de Informática

---



---

## Visualización de datos de comunidades colaborativas

---

Trabajo de fin de grado del Grado en Ingeniería  
Informática

Curso 2015/2016

*Autores:*

Claudia Gil Navarro

Alejandro Pequeño Pulleiro

Alejandro Jesús del Valle Silva

*Directores de proyecto:*

Javier Arroyo Gallardo

Samer Hassan Collado



# Autorización de difusión y uso

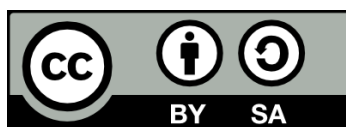
Autorizo a la Universidad Complutense de Madrid a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor, tanto la propia memoria, como el código, la documentación y/o el software desarrollado.

Claudia Gil Navarro

Alejandro Pequeño Pulleiro

Alejandro del Valle Silva

Madrid, Junio 2016



Copyleft by Claudia Gil Navarro, Alejandro Pequeño Pulleiro and Alejandro Jesús del Valle Silva, released under the license Creative Commons Attribution Share-Alike International 4.0 available at:

[https://creativecommons.org/licenses/by-sa/4.0/deed.es\\_ES](https://creativecommons.org/licenses/by-sa/4.0/deed.es_ES)



# Índice general

<b>RESUMEN .....</b>	<b>13</b>
<b>ABSTRACT .....</b>	<b>14</b>
<b>CAPÍTULO 1.....</b>	<b>16</b>
<b>INTRODUCCIÓN .....</b>	<b>16</b>
1.1. ANTECEDENTES .....	16
1.1.1. Marco teórico: la era de la visualización.....	16
1.1.2. Motivación detrás de ChartsUp .....	17
1.2. OBJETIVOS DEL PROYECTO .....	18
1.3. ESTRUCTURA DEL DOCUMENTO .....	18
<b>CHAPTER 1 .....</b>	<b>20</b>
<b>INTRODUCTION .....</b>	<b>20</b>
1.1. ANTECEDENT .....	20
1.1.1. Theoretical scenario: the era of visualization .....	20
1.1.2. Motivation behind ChartsUp .....	21
1.2. GOALS .....	22
1.3. STRUCTURE OF THE REPORT .....	22
<b>CAPÍTULO 2.....</b>	<b>25</b>
<b>ESTADO DEL ARTE.....</b>	<b>25</b>
2.1. PRODUCCIÓN COLABORATIVA .....	25
2.2. COMUNIDADES COLABORATIVAS .....	27
2.2.1. Wikia .....	29
2.3. ACCESO A DATOS .....	31
Open Collaboration Data Factories .....	33
Open Referral .....	33
DERP.....	33
2.4. HERRAMIENTAS DE VISUALIZACIÓN Y ANÁLISIS DE INFORMACIÓN .....	34
2.4.1. Herramientas de Análisis Estadístico .....	34
WAM de Wikia .....	34
Quantcast.....	35
Alexa Ranking .....	36
Bitergia Cauldron .....	36
2.4.2. Servicios de visualización .....	38
2.4.2.1. Aplicaciones de visualización.....	38
Google Fusion Tables .....	38
CartoDB.....	39
Tableau Software .....	39
2.4.2.2. Librerías y frameworks .....	40
D3.js .....	41
Google Chart Tools.....	41
Chart.js.....	42
Echarts.js.....	42
2.4.3. Herramientas para el Análisis de Redes.....	43
NodeXL.....	43
Gephi.....	43

2.5. CONCLUSIONES.....	44
<b>CAPÍTULO 3.....</b>	<b>46</b>
<b>METODOLOGÍA DEL PROYECTO .....</b>	<b>46</b>
3.1. PLAN DE TRABAJO .....	46
3.2. ENFOQUE DE SOFTWARE LIBRE.....	47
3.3. DISEÑO DE LA APLICACIÓN WEB .....	47
3.3.1. <i>Brainstorming</i> .....	48
3.3.2. <i>Diseño Guiado por Objetivos</i> .....	48
<b>CAPÍTULO 4.....</b>	<b>51</b>
<b>TECNOLOGÍAS DEL PROYECTO .....</b>	<b>51</b>
4.1. HTML.....	52
4.2. CSS .....	53
4.3. JAVASCRIPT .....	53
4.4. PHP.....	54
4.5. SQL .....	54
4.6. ECHARTS .....	54
4.7. JQUERY.....	55
4.8. BOOTSTRAP .....	55
4.9. MYBALSAMIQ.....	56
4.10. JUSTINMIND.....	56
4.11. MYSQL .....	56
4.12. PHPMYADMIN .....	57
4.13. XAMPP .....	57
4.14. SUBLIME TEXT .....	58
4.15. WRIKE .....	58
4.16. GITHUB .....	59
<b>CAPÍTULO 5.....</b>	<b>61</b>
<b>DISEÑO DE LA APLICACIÓN .....</b>	<b>61</b>
5.1. INTRODUCCIÓN.....	61
5.2. DISEÑO GUIADO POR OBJETIVOS .....	61
5.2.1. <i>Investigación</i> .....	61
5.2.2. <i>Modelado</i> .....	63
5.2.2.1. Identificación de categorías de usuarios .....	63
5.2.2.2. Procesamiento de los datos y esqueletos .....	63
5.2.2.3. Priorización de esqueletos .....	65
5.2.3. <i>Definición de requisitos</i> .....	65
5.2.4. <i>Framework de diseño</i> .....	66
5.2.4.1. Definir el factor de forma, la postura y los métodos de entrada .....	67
5.2.4.2. Definir los elementos y datos funcionales.....	67
5.2.4.3. Determinar los grupos funcionales .....	68
5.2.4.4. Boceto del framework de interacción .....	69
Escenarios Key Path y validación.....	69
Principios de diseño .....	75
5.3. DISEÑO DE PANELES .....	77
5.3.1. <i>Introducción</i> .....	77
5.3.2. <i>Paneles de ChartsUp</i> .....	79
5.3.2.1. Comparador de wikis.....	80

5.3.2.2. Perfil de una wiki .....	82
5.3.2.3. Estadísticas de wikis .....	83
5.3.2.4. Estadísticas de usuarios.....	85
5.3.2.5. Perfil de usuario .....	88
<b>CAPÍTULO 6.....</b>	<b>91</b>
<b>ARQUITECTURA DEL PROYECTO.....</b>	<b>91</b>
6.1. RECOPIRAR INFORMACIÓN .....	91
6.1.1. <i>Wikia Content API</i> .....	91
6.1.1.1. Problema de la latencia .....	94
6.1.2. <i>Database dumps de Wikia</i> .....	95
6.1.3. <i>PHP Simple HTML DOM Parser</i> .....	96
6.1.3.1. Scripts.....	98
6.2. BASE DE DATOS .....	103
6.2.1. Wikis.....	104
6.2.2. Aportaciones .....	105
6.2.3. Logros.....	106
6.2.4. Comparaciones.....	106
6.2.5. Novedades_app.....	107
6.2.6. Preguntas .....	107
6.2.7. Ideas_futuro .....	108
6.3. DISEÑO WEB .....	109
6.4. SERVIDOR WEB .....	112
<b>CAPÍTULO 7.....</b>	<b>114</b>
<b>EVALUACIÓN CON USUARIOS .....</b>	<b>114</b>
7.1. OBJETIVOS DE LA EVALUACIÓN.....	114
7.2. CARACTERÍSTICAS DE LA EVALUACIÓN.....	114
7.3. RESULTADOS DE LA EVALUACIÓN .....	116
7.3.1. <i>Interfaz gráfica</i> .....	116
7.3.2. <i>Calidad de la información</i> .....	118
<b>CAPÍTULO 8.....</b>	<b>121</b>
<b>FUNCIONALIDAD FINAL DE CHARTSUP.....</b>	<b>121</b>
8.1.1. Welcome page.....	121
8.1.2. Compare.....	122
8.1.3. Search.....	123
8.1.4. Perfil de una wiki .....	124
8.1.5. Wikis Statistics.....	125
8.1.6. Users Statistics .....	128
8.1.7. Perfil de usuario .....	130
8.1.8. About us .....	131
<b>CAPÍTULO 9.....</b>	<b>134</b>
<b>RESULTADOS.....</b>	<b>134</b>
9.1. DISCUSIÓN DE LOS RESULTADOS .....	134
9.2. REPARTO DE TRABAJO .....	135
9.2.1. Alejandro Pequeño Pulleiro.....	135
9.2.2. Claudia Gil Navarro.....	136
9.2.3. Alejandro del Valle Silva .....	137

<b>CAPÍTULO 10.....</b>	<b>139</b>
<b>CONCLUSIONES Y TRABAJO FUTURO .....</b>	<b>139</b>
10.1. CONCLUSIONES .....	139
10.2. TRABAJO FUTURO .....	141
<b>CHAPTER 10 .....</b>	<b>143</b>
<b>CONCLUSIONS AND FUTURE WORK .....</b>	<b>143</b>
10.1. CONCLUSIONS .....	143
10.2. FUTURE WORK.....	144
<b>BIBLIOGRAFÍA .....</b>	<b>148</b>



# Índice de figuras

FIGURA 1. ECONOMÍA COLABORATIVA .....	25
FIGURA 2. DONACIONES A WIKIMEDIA ENTRE EL AÑO 2013 Y 2014.....	27
FIGURA 3. EVOLUCIÓN DEL NÚMERO DE ARTÍCULOS ACTIVOS EN FUNCIÓN DEL IDIOMA .....	28
FIGURA 4. PERFIL DE USUARIOS QUE MÁS VISITAS REALIZAN A WIKIA .....	31
FIGURA 5. DISTRIBUCIÓN DEL PRESUPUESTO EN LA COMUNIDAD DE MADRID PARA EL AÑO 2015.....	32
FIGURA 6. CLASIFICACIÓN WAM PARA LENGUAJE ESPAÑOL A 2 DE JUNIO DE 2016.....	35
FIGURA 7. <i>COMMITTS</i> A LO LARGO DEL TIEMPO PARA ANDROID DE CAULDRON.....	36
FIGURA 8. PANEL DE <i>PULL REQUESTS</i> PARA GIT DE CAULDRON.....	37
FIGURA 9. GOOGLE FUSION TABLES: THE SUBMARINE CABLE MAP .....	38
FIGURA 10. CARTODB: IMPACTO DEL FLUJO DE INMIGRANTES EN POBLACIÓN Y MUERTES (2000-2015).....	39
FIGURA 11. TABLEAU: EMISIONES MUNDIALES DE CO2 POR GAS Y PETRÓLEO .....	40
FIGURA 12. D3.JS: FLUCTUACIÓN DINÁMICA DE LOS INGRESOS PER CÁPITA (X), LA ESPERANZA DE VIDA (Y) Y POBLACIÓN (RADIO) DE 180 NACIONES EN LOS ÚLTIMOS 209 AÑOS.....	41
FIGURA 13. GEPHI: TRAYECTORIA DE LOS REFUGIADOS SEGÚN EL INFORME DE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES.....	43
FIGURA 14. NUBE DE TECNOLOGÍAS DEL PROYECTO .....	51
FIGURA 15. BALANCE DE TECNOLOGÍAS JAVASCRIPT, HTML Y CSS EN LA WEB.....	53
FIGURA 16. EJEMPLO DE GRÁFICO DE BARRAS DE ECHARTS.....	55
FIGURA 17. PHPMYADMIN .....	57
FIGURA 18. PANEL DE CONTROL DE XAMPP.....	58
FIGURA 19. PANEL DE GESTIÓN DE PROYECTOS DE WRIKE .....	59
FIGURA 20. FICHAS FICTICIAS DE PERFILES DE USUARIOS DE CHARTSUP .....	65
FIGURA 21. BOCETO DE LA PÁGINA DE BIENVENIDA DE LA APLICACIÓN.....	70
FIGURA 22. BOCETO DEL ACCESO AL COMPARADOR.....	70
FIGURA 23. BOCETO DE LA PÁGINA DEL COMPARADOR.....	71
FIGURA 24. BOCETO DE LA PÁGINA DE GLOBAL USERS .....	72
FIGURA 25. BOCETO DEL PERFIL DE UN USUARIO DE WIKIA .....	72
FIGURA 26. BOCETO DEL ACCESO AL BUSCADOR DE WIKIS .....	73
FIGURA 27. BOCETO DE LA TABLA DE BÚSQUEDA DE WIKIS .....	74
FIGURA 28. BOCETO DEL PERFIL DE UNA WIKI .....	74
FIGURA 29. ENLACES OCULTOS/VISIBLES DEL MENÚ PRINCIPAL .....	76
FIGURA 30. ESTADO LOADING DE PÁGINAS DE CHARTSUP .....	77
FIGURA 31. EJEMPLO DE MAL USO DE UN GRÁFICO DE BARRAS .....	78
FIGURA 32. EJEMPLO DE GRÁFICOS SIN DEFINICIÓN EN SUS EJES X E Y .....	78
FIGURA 33. EJEMPLO DE GRÁFICOS TIPO TARTA MAL DISEÑADOS.....	79
FIGURA 34. COMPARADOR DE PÁGINAS POR WIKIS.....	80
FIGURA 35. USUARIOS REGISTRADOS Y EDICIONES DEL COMPARADOR DE WIKIS.....	81
FIGURA 36. TIPOS DE VISUALIZACIÓN PARA EL TIPO DE USUARIO DE CADA WIKI COMPARADA .....	81
FIGURA 37. VISUALIZACIONES PARA LA ACTIVIDAD DE USUARIOS DEL COMPARADOR .....	82
FIGURA 38. PARTICIPACIÓN EN EL ÚLTIMO MES PARA LA WIKI .....	82
FIGURA 39. ESTADÍSTICAS DE UNA WIKI PARA LA CATEGORÍA EDICIONES .....	83
FIGURA 40. GRÁFICO DE ANILLO DE LA SECCIÓN TOP ACTIVE USERS .....	83
FIGURA 41. TIPOS DE CORRELACIÓN .....	84
FIGURA 42. SCATTER GRAPH PARA EDICIONES (EJE X) Y PÁGINAS (EJE Y).....	84
FIGURA 43. EVOLUCIÓN DEL NÚMERO DE WIKIS EDITADAS DESDE 2005 HASTA 2016.....	85
FIGURA 44. TOP USUARIOS MÁS ACTIVOS POR NÚMERO DE EDICIONES.....	86
FIGURA 45. ACTIVIDAD DE LOS USUARIOS (PROPORCIÓN) POR NÚMERO DE EDICIONES .....	86

FIGURA 46. TOP USUARIOS MÁS ACTIVOS POR NÚMERO DE WIKIS EN LAS QUE PARTICIPAN .....	87
FIGURA 47. ACTIVIDAD DE LOS USUARIOS (PROPORCIÓN) POR NÚMERO DE WIKIS.....	88
FIGURA 48. NÚMERO DE EDICIONES EN CADA WIKI EN LA QUE PARTICIPA EL USUARIO.....	88
FIGURA 49. PARTICIPACIÓN DE UN USUARIO, POR AÑO (IZQ) Y EN PROPORCIÓN (DCHA) .....	89
FIGURA 50. RESULTADOS DE CONSULTA PARA LA LISTA DE WIKIS CON SUS PRUNTUACIÓN EN EL WAM.....	92
FIGURA 51. INFORMACIÓN SOBRE ADMINISTRADORES DEVUELTA POR LA API DE WIKIA .....	93
FIGURA 52. RESULTADOS DE LA CONSULTA PARA LA ACTIVIDAD RECIENTE DE UNA WIKI.....	93
FIGURA 53. TIEMPO DE LLAMADAS A LA API DE WIKIA .....	94
FIGURA 54. EJEMPLO DE DUMP DE LA WIKI MUPPET .....	96
FIGURA 55. PHP SIMPLE HTML DOM PARSER DOCUMENTACIÓN .....	96
FIGURA 56. CÓDIGO DE LA PÁGINA "LISTA DE COMUNIDADES" DE WIKIA.....	97
FIGURA 57. CÓDIGO DE BÚSQUEDA DE TODAS LAS COMUNIDADES DE WIKIA.....	98
FIGURA 58. NIVELES DE ACCESO A LAS APORTACIONES DE USUARIOS PARA CADA UNA DE LAS WIKIS.....	99
FIGURA 59. PÁGINA DE ACTIVIDAD DE LA WIKI POKEMON .....	99
FIGURA 60. FORMATO QUE SIGUE WIKIACTIVITY QUE PROPORCIONA LAS APORTACIONES DE USUARIOS.....	100
FIGURA 61. SECCIÓN DEL CÓDIGO DE PARSEO DE WIKIACTIVITY PARA OBTENER APORTACIONES .....	100
FIGURA 62. CÓDIGO PARA DESCARTAR USUARIOS NO REGISTRADOS.....	100
FIGURA 63. EJEMPLO DE PÁGINA DE ESTADÍSTICAS DE UNA WIKI.....	102
FIGURA 64. ESQUEMA DE RELACIÓN DE LA BASE DE DATOS.....	103
FIGURA 65. TRANSICIÓN DE UNA VISUALIZACIÓN A OTRA DE ECHARTS .....	109
FIGURA 66. DEFINICIÓN DE LA ESTRUCTURA DE DOS PANELES DE ECHARTS.....	110
FIGURA 67. CÓDIGO DE DEFINICIÓN DE LAS OPCIONES DE UN PANEL (1) .....	111
FIGURA 68. CÓDIGO DE DEFINICIÓN DE LAS OPCIONES DE UN PANEL (2) .....	111
FIGURA 69. ERRORES EN EL TEXTO DE ALGUNOS GRÁFICOS.....	117
FIGURA 70. GRÁFICO QUE MUESTRA A LA VEZ CONTENIDO, FICHEROS, EDICIONES Y PÁGINAS DE VARIAS WIKIS .....	119
FIGURA 71. PANTALLA DE BIENVENIDA DE LA APLICACIÓN .....	121
FIGURA 72. ACCESO AL COMPARADOR.....	122
FIGURA 73. COMPARADOR DE 3 WIKIS.....	122
FIGURA 74. ACTIVIDAD DE USUARIOS Y TABLA DE LÍDERES DE UNA DE LAS WIKIS DEL COMPARADOR.....	123
FIGURA 75. BÚSQUEDA ALFABÉTICA DE WIKIS.....	124
FIGURA 76. INFORMACIÓN GENERAL DEL PERFIL DE UNA WIKI.....	124
FIGURA 77. LISTA DE USUARIOS (IZDA) Y PARTICIPACIÓN (DCHA) EN LA WIKI.....	125
FIGURA 78. MEDIA DE EDICIONES Y FICHEROS PARA ESA WIKI .....	125
FIGURA 79. TOP ADMINISTRADORES, CON LA VISUALIZACIÓN DEL RESTO DE WIKIS DESMARCADA .....	126
FIGURA 80. LAS 10 WIKIS CON MÁS ADMINISTRADORES.....	127
FIGURA 81. PERSONALIZACIÓN DE LOS VALORES DEL SCATTER GRAPH.....	127
FIGURA 82. GRÁFICO DE DISPERSIÓN PARA EDICIONES (EJE X) Y PÁGINAS (EJE Y) .....	127
FIGURA 83. EVOLUCIÓN DE LA PARTICIPACIÓN EN WIKIA.....	128
FIGURA 84. USUARIOS CON MAYOR NÚMERO DE EDICIONES.....	128
FIGURA 85. ACTIVIDAD (PROPORCIÓN) DE LOS USUARIOS (IZDA) Y LOS 500 USUARIOS MÁS ACTIVOS (DCHA), .....	129
FIGURA 86. ACTIVIDAD DE LOS USUARIOS EN FUNCIÓN DEL NÚMERO DE WIKIS EN LAS QUE PARTICIPAN .....	129
FIGURA 87. CABECERA DEL PERFIL DE UN USUARIO .....	130
FIGURA 88. PARTICIPACIÓN DE UN USUARIO POR AÑO (IZDA) Y EN PROPORCIÓN (DCHA).....	130
FIGURA 89. TABLA CON LOS ÚLTIMOS LOGROS CONSEGUIDOS POR EL USUARIO .....	131
FIGURA 90. SECCIÓN ABOUT US DE CHARTSUP.....	131
FIGURA 91. EVOLUCIÓN DEL DESARROLLO DE CHARTSUP .....	132
FIGURA 92. CÓDIGO SCRAPY QUE SIGUE LOS ENLACES A LAS PREGUNTAS MÁS VOTADAS DE STACKOVERFLOW Y "ARAÑA"	
DATOS DE CADA PÁGINA .....	142

# Índice de tablas

TABLA 1. TECNOLOGÍAS DEL PROYECTO .....	52
TABLA 2. WIKIS .....	104
TABLA 3. APORTACIONES .....	105
TABLA 4. LOGROS .....	106
TABLA 5. COMPARACIONES .....	106
TABLA 6. NOVEDADES DE LA APLICACIÓN .....	107
TABLA 7. PREGUNTAS.....	108
TABLA 8. IDEAS PARA DESARROLLOS FUTUROS .....	108



# RESUMEN

Las comunidades colaborativas, donde grandes cantidades de personas colaboran para la producción de recursos compartidos (e.g. Github, Wikipedia, OpenStreetMap, Arduino, StackOverflow) están extendiéndose progresivamente a multitud de campos. No obstante, es complicado comprender cómo funcionan y evolucionan. ¿Qué tipos de usuarios son más activos en Wikia? ¿Cómo ha evolucionado el número de wikis activas en los últimos años? ¿Qué perfil de actividad presentan la mayor parte de colaboradores de Wikia? ¿Son más activos los hombres o las mujeres en la Wikipedia? En los proyectos de Github, ¿el esfuerzo de programación (y frecuencia de *commits*) se distribuye de forma homogénea a lo largo del tiempo o suele estar concentrado? Estas comunidades, típicamente online, dejan registrada su actividad en grandes bases de datos, muchas de ellas disponibles públicamente. Sin embargo, el ciudadano de a pie no tiene ni las herramientas ni el conocimiento necesario para sacar conclusiones de esos datos. En este TFG desarrollamos una herramienta de análisis exploratorio y visualización de datos de la plataforma Wikia, sitio web colaborativo que permite la creación, edición y modificación del contenido y estructura de miles de páginas web de tipo enciclopedia basadas en la tecnología wiki.

Nuestro objetivo es que esta aplicación web sea usable por cualquiera y que no requiera que el usuario sea un experto en Big Data para poder visualizar las gráficas de evolución o distribuciones del comportamiento interno de la comunidad, pudiendo modificar algunos de sus parámetros y visualizando cómo cambian.

Como resultado de este trabajo se ha desarrollado una primera versión de la aplicación disponible en GitHub<sup>1</sup> y en <http://chartsup.esy.es/>

**Palabras clave:** Comunidades colaborativas, visualización, análisis, Wikia, producción colaborativa, aplicación web.

---

<sup>1</sup> <https://github.com/Grasia/ChartsUp>

# ABSTRACT

Collaborative communities, where large numbers of people work together and share the common goods they produce (Github, Wikipedia, OpenStreetMap, Arduino, StackOverflow), are gradually spreading to many different areas. However, it is difficult to understand how they work and evolve. What types of users are more active on Wikia? How has evolved the number of active wikis in recent years? What activity profile present most of Wikia collaborators? Are more active men or women in Wikipedia? Programming effort (and frequency of commits) is distributed evenly or is it usually focused in Github projects? These typically online communities register their activity in large databases, many of them publicly available. Even so, the ordinary citizen has neither the tools nor the knowledge to draw conclusions from these data. In this project we develop a tool for exploratory data analysis and visualization of the Wikia platform, a collaborative website that allows the creation, editing and modification of the content and structure of thousands of encyclopedia type web pages based on wiki technology.

We aim to produce a web tool usable by anyone and that doesn't require the user to view display graphs or distributions of the internal behavior of the community, being able to modify some of its parameters and visualizing how it change.

The first versión of the web application is available at GitHub <sup>2</sup> and at <http://chartsup.esy.es/>

**Keywords:** Collaborative communities, visualization, analysis, Wikia, collaborative production, web application.

---

<sup>2</sup> <https://github.com/Grasia/ChartsUp>



# CAPÍTULO 1

## INTRODUCCIÓN

A continuación se expone el contexto en el que se sitúa ChartsUp, así como algunas observaciones acerca del porqué de nuestra aplicación. Después se detallan los objetivos de este proyecto y la estructura del documento.

### 1.1. ANTECEDENTES

En los siguientes apartados abriremos debate en torno a la creciente importancia de la visualización de datos en la actualidad, y su relación con el desarrollo de nuestra herramienta.

#### 1.1.1. MARCO TEÓRICO: LA ERA DE LA VISUALIZACIÓN

Como sabemos, somos una sociedad por Internet con abundancia de datos, incluso saturación de datos. Se calcula que el 90% de todos los datos que existen en la actualidad se han generado en los dos últimos años [1]. Esa abundancia hace que muchas veces resulte complejo diferenciar lo que es importante de lo que no. Sin embargo, y contrariamente a lo que pueda parecer a primera vista, para que esa información facilite el acceso al conocimiento es necesario que ese gran volumen de datos se organice y transforme en información comprensible. Es el acceso a esos datos y la capacidad para analizarlos y relacionarlos lo que realmente les da valor. Si a esto le sumamos que nuestra capacidad de retención de información es bastante limitada, necesitamos algo que nos sintetice la información.

De esto se encarga, entre otras disciplinas, la visualización de datos, parte fundamental de este proyecto. La visualización de datos es el **proceso de interpretación, contrastación y comparación de datos que permite adquirir un conocimiento en profundidad de los mismos de manera que se transformen en información comprensible para el usuario.**

El trabajo con datos en crudo es complejo, y debemos trabajar con ellos para darles estructura. Muchos de los tutoriales de visualización de datos que podemos encontrar en internet muestran unos datasets ordenados y limpios. Sin embargo, en la práctica no ocurre así. La mayoría de conjuntos de datos que nos encontramos en el mundo real requiere que dediquemos cerca del 80% de nuestro tiempo [2] en la creación de una



visualización adecuada, en la exploración, adquisición, carga y posterior transformación de dichos datos, aun contando con herramientas que automaticen la limpieza de los mismos.

Es decir, hacer una visualización de un conjunto voluminoso de datos no es una tarea trivial. Como afirma Stephen Few, las habilidades requeridas para visualizar la información de forma efectiva no son intuitivas y dependen en gran medida de principios que deben ser aprendidos [3]. Además de analizar la información, saber cómo interpretarla, contrastarla con otra y estudiarla, hay que saber **comunicarla** e incluirla dentro de un contexto adecuado. No obstante, es necesario recordar que **la visualización no es el análisis**. El análisis de datos es el **proceso de examinar, limpiar y transformar datos con el objetivo de resaltar información útil y sacar conclusiones**.

Es fundamental insistir en que la visualización es una herramienta que proporciona una ayuda al análisis, y que en ningún caso es una sustitución del trabajo analítico. Tampoco es un sustituto de las estadísticas: los gráficos de barras, por ejemplo, pueden resultar útiles para resaltar las diferencias o correlaciones entre los datos, pero para sacar conclusiones de forma fiable sobre percepciones muy a menudo requerimos de un enfoque estadístico más riguroso. Si estamos un poco avezados en programación, disponemos de diversas herramientas, como librerías de JavaScript estilo Raphaël o Paper.js que, aprovechando funcionalidades tipo canvas, nos facilitan mucho la tarea de producir visualizaciones de datos complejas, hasta hace poco área exclusiva de programadores avanzados. Esto estrecha aún más la relación entre una disciplina como la informática y el auge de este tipo de herramientas, potenciando sus características a medida que evolucionan las tecnologías.

La visualización y el análisis, además, son herramientas que se pueden y, de hecho, se utilizan, en múltiples campos. Desde el marketing digital hasta estudios científicos, en los que los gráficos comunican tendencias y datos complejos que solo con números, tablas y texto sería más difícil comprender la información que hay detrás.

### 1.1.2. MOTIVACIÓN DETRÁS DE CHARTSUP

La idea a desarrollar se basa en la necesidad de dar acceso a una determinada información que, aun estando disponible y al alcance de cualquiera, no resulta fácilmente entendible. Esta podría, por tanto, considerarse una herramienta que facilitara la comprensión de unos datos que, en sí mismos y sin trabajar, no permiten extraer conclusiones de una manera simple.

La herramienta debe ofrecer una interfaz atractiva que permite que cualquiera pueda analizar el comportamiento de una plataforma colaborativa sin necesidad de realizar primero una limpieza o estructuración de la información para llegar a entenderla. Así, **cualquier usuario tendría a su disposición a una información procesada de tal forma que no suponga un esfuerzo acceder a ella**. A su vez, esa presentación debe ser

adecuada para el tipo de información que se quiere mostrar. Supondría dar un paso atrás y dificultar su comprensión utilizar unos gráficos que confundieran al usuario.

Por tanto, desde un primer punto de vista, la aplicación está dividida en dos partes. Por un lado se depuran y sintetizan los datos en bruto, y por otro se estructura y adapta el tipo de visualización al dato para sacar el máximo partido a la herramienta.

## 1.2. OBJETIVOS DEL PROYECTO

El objetivo general del proyecto es **definir una herramienta que permita visualizar de forma flexible la información que se encuentra almacenada de diferentes formas en las comunidades de una de las plataformas colaborativas de mayor actividad, Wikia, así como consultar datos concretos** para poder hacer análisis sobre ellos. Se trata de que la aplicación permita tanto curiosear a usuarios casuales como facilitar la extracción de datos a sociólogos, periodistas o analistas de datos, y facilitar su trabajo.

En los siguientes puntos detallamos este objetivo general:

- 1ª parte: Realizar análisis de los principales datos de las comunidades de Wikia
  - Aprender sobre visualización de información web y análisis de comunidades colaborativas
  - Visibilizar los principales contribuyentes en cada comunidad Wikia
- 2ª parte: Desarrollar una aplicación web que recopile la información obtenida y permita al usuario interactuar con ella
  - Proporcionar información sencilla de comprender a usuarios no experimentados
  - Proporcionar a usuarios más experimentados una herramienta que permita extraer información más detallada y cruzar datos entre comunidades
  - Proporcionar un proyecto de software libre modular para que otros usuarios interesados lo amplíen en el futuro

## 1.3. ESTRUCTURA DEL DOCUMENTO

Esta memoria recoge los aspectos fundamentales del desarrollo de este proyecto, desde la parte más puramente de investigación hasta los aspectos técnicos. A pesar de desarrollar una herramienta web, no obstante, este documento solo describe a grandes rasgos el funcionamiento de la misma y en ningún caso sirve como manual o tutorial. La estructura seguida a lo largo de la memoria se basa en capítulos con sus

correspondientes subapartados, en los que se describirá de la manera más detallada posible las distintas fases y características del desarrollo del proyecto:

- **Capítulo 1. Introducción.** Incluye un breve análisis sobre la idea detrás de ChartsUp, así como los objetivos del proyecto y estructura de este documento.
- **Capítulo 2. Estado del arte.** En este capítulo se intenta proporcionar una visión general de las técnicas y herramientas relacionadas con la visualización y el análisis de información.
- **Capítulo 3. Metodología del proyecto.** Se explica brevemente la metodología seguida durante el desarrollo de este Trabajo de Fin de Grado.
- **Capítulo 4. Tecnologías del proyecto.** Incluye una pequeña descripción de las tecnologías principales utilizadas en el desarrollo, explicando en qué consisten y por qué se han elegido estas y no otras.
- **Capítulo 5. Diseño de la aplicación.** En este capítulo se incluyen todas las fases incluidas en el Diseño Guiado por Objetivos que sirve de guía para este proyecto, incluidos los prototipos de la aplicación y el detalle de los paneles.
- **Capítulo 6. Arquitectura del proyecto.** Define las características más técnicas de la aplicación, desde la estructura de la base de datos hasta el proceso de captura de información. También expone la evaluación con usuarios y las dos iteraciones en las que se ha basado el desarrollo de la herramienta.
- **Capítulo 7. Evaluación con usuarios.** Incluye el proceso seguido durante la evaluación de la aplicación con usuarios reales: objetivos, características y conclusiones obtenidas.
- **Capítulo 8. Funcionalidad final de ChartsUp.** Se describe la funcionalidad detallada de la aplicación.
- **Capítulo 9. Resultados.** Se plantea una discusión sobre los resultados obtenidos y se especifica el reparto de tareas.
- **Capítulo 10. Conclusiones y trabajo futuro.** Se exponen las conclusiones obtenidas tras completar este proyecto y se establecen los pasos a seguir en caso de querer extender la funcionalidad de la aplicación.

Los capítulos 1 y 10 han sido traducidos al inglés, como se solicita en las indicaciones para la redacción de la memoria de TFG.

# CHAPTER 1

## INTRODUCTION

In the next paragraphs, the context in which ChartsUp is located, as well as some observations about the reasons behind our application are exposed, followed by the goals of this project and the structure of the paper.

### 1.1. ANTECEDENT

In the following sections we will open discussion on the growing importance of data visualization nowadays, and its relation to the development of our tool.

#### 1.1.1. THEORETICAL SCENARIO: THE ERA OF VISUALIZATION

As we know, we are an Internet society with an abundance of data, even data saturation. It is estimated that 90% of all data that exist today have been generated in the last two years [1]. That amount of information, especially visual, means that it is sometimes very difficult to spot the difference between what is important and what is not. However, contrary to what it may seem at first glance, it is necessary for large volume of data to be organized and transformed into understandable information in order to allow access to knowledge.

The access to that data and the ability to analyze it properly is what really offer added value. If we add that our ability to retain information is rather limited, we need something that synthesizes information.

Data visualization aims to produce this kind of effect, among other disciplines. Data visualization is the **process of interpretation, contrasting and comparing data that allows to acquire in-depth knowledge of them so that they become understandable to the user.**

Working with raw data is complex, and we must work to give it structure. Many of the data visualization tutorials we find on the internet show a neat and clean datasets. Nevertheless, this does not happen in the real world. Most datasets that we find requires us to devote about 80% of our time [2] in creating adequate visualization, in the

exploration, acquisition, freight and further processing of the data, even with data cleaning tools that help us.

That is to say, the visualization of large datasets is not a trivial task. As Stephen Few explains, the skills required to display information effectively are not intuitive and depend largely on principles that must be learned [3]. In addition to analyzing the information, know how to interpret it, contrast and study it, one must know how to communicate and include it in a proper context. However, we must not forget that **visualization is not analysis**. Data analysis is the **process of examining, cleaning and transforming data with the aim of highlighting useful information and draw conclusions**.

It is essential to stress that visualization is a tool that provides help to the analysis, and in no case is a replacement of the analytical work. Nor is it a substitute for the statistics: bar charts, for example, can be useful to highlight the differences or correlations between data, but to draw conclusions reliably on perceptions often require a more rigorous statistical approach. If we are a little seasoned programming, we have various tools such as JavaScript libraries, like Raphaël or Paper.js, that make it much easier to produce visualizations of complex data, exclusive area of advanced programmers until recently. This narrows further the relationship between a discipline such as computer science and the rise of these tools, enhancing its features as technologies evolve.

Visualization and analysis are also tools that can and, in fact, are used in different areas. From digital marketing to scientific studies, in which communicate complex data only with numbers and text would be difficult to understand.

### 1.1.2. MOTIVATION BEHIND CHARTSUP

The aim is to provide access to certain information which, although available and accessible to anyone, is not easily understandable. This could, therefore, be considered a tool to facilitate the understanding of data that, untreated, does not allow to draw conclusions in a simple way.

The tool offers an attractive interface that allows anyone to analyze the behavior of a collaborative platform without first performing cleaning or structuring information to get to understand it. Thus, **any user would have access to an elaborate information, so it presents no effort to get it**. In turn, this presentation must be appropriate for the type of information you want to display. It would take a step back and make it difficult to understand to use graphics and charts that confuse the user.

Therefore, from a first point of view, the application is divided into two parts. On the one hand, it synthesizes and cleans information, and on the other hand, it gives structure to the data and adapts the display to get the most out of the tool.

## 1.2. GOALS

The main goal of the project is to **define a tool to flexibly display the information that is stored in different ways in the communities of one of the busiest collaborative platforms, Wikia, and extract specific data** to perform analysis on it. This application will be able to allow both casual users to browse, and to make the data extraction easier to sociologists, journalists or data analysts, and facilitate their work.

The following sections detail this overall objective:

- 1<sup>st</sup> part: Conduct analysis of key data Wikia communities
  - Learn about displaying web information and analysis of collaborative communities
  - Visualizing the main contributors in each Wikia community
- 2<sup>nd</sup> part: Develop a web application that collects the information obtained and allow the user to interact with it
  - Provide simple and easy to understand information to inexperienced users
  - Provide a web tool to extract detailed information to more experienced users, as well as to allow cross data between Wikia communities
  - Provide a modular free software project to other interested users

## 1.3. STRUCTURE OF THE REPORT

This document includes the fundamental aspects of the development of the project, from the research part to the more purely technical content. While developing a web tool, however, this document only outlines the application and in any case serves as a manual or tutorial. The structure followed over this report is based on chapters and subsections. In each one of them will be described in as much detail as possible the different phases and characteristics of the project.

- **Chapter 1. Introduction.** It includes an analysis about the reasons behind ChartsUp, as well as the goals of the project and structure of this document.
- **Chapter 2. State of art.** This chapter attempts to provide an overview of the techniques and related visualization and data analysis tools.
- **Chapter 3. Methodology.** This chapter summarizes the methodology used during the development of this project.
- **Chapter 4. Technologies.** It includes a brief description of the main tools and technologies used in the development, explaining what they are and why they have been chosen.
- **Chapter 5. Design.** It contains the phases included in the Goal-Directed Design that serves as a guide for this project, including the prototypes and the dashboard design.

- **Chapter 6. Architecture.** It defines the most technical aspects of the project, as well as the two main iterations in which the development has been divided.
- **Chapter 7. Users experience evaluation.** This chapter contains the whole user experience evaluation: goals, characteristics and final conclusions.
- **Chapter 8. ChartsUp's final functionality.** It describes the functionality of the web tool.
- **Chapter 9. Results.** It contains the discussion of the results obtained and the task distribution.
- **Chapter 10. Conclusions and future work.** The conclusions obtained after completing the project are discussed in this chapter, as well as the steps to follow in case you want to extend the functionality of the web tool.





# CAPÍTULO 2

## ESTADO DEL ARTE

En este capítulo se hablará tanto de la producción colaborativa y plataformas basadas en esta iniciativa, como de algunas herramientas disponibles actualmente relacionadas con la visualización y el análisis de datos. Todo ello con el objetivo de establecer una referencia que ayude al lector a situar este proyecto en el contexto adecuado.

### 2.1. PRODUCCIÓN COLABORATIVA

Hoy en día es muy común escuchar sobre la producción colaborativa. Dado que este proyecto se centra en la visualización de información procedente de una comunidad colaborativa de contenido cultural, creemos que es importante presentar una serie de conceptos relacionados con la economía y producción colaborativa, centrándonos en bienes digitales.

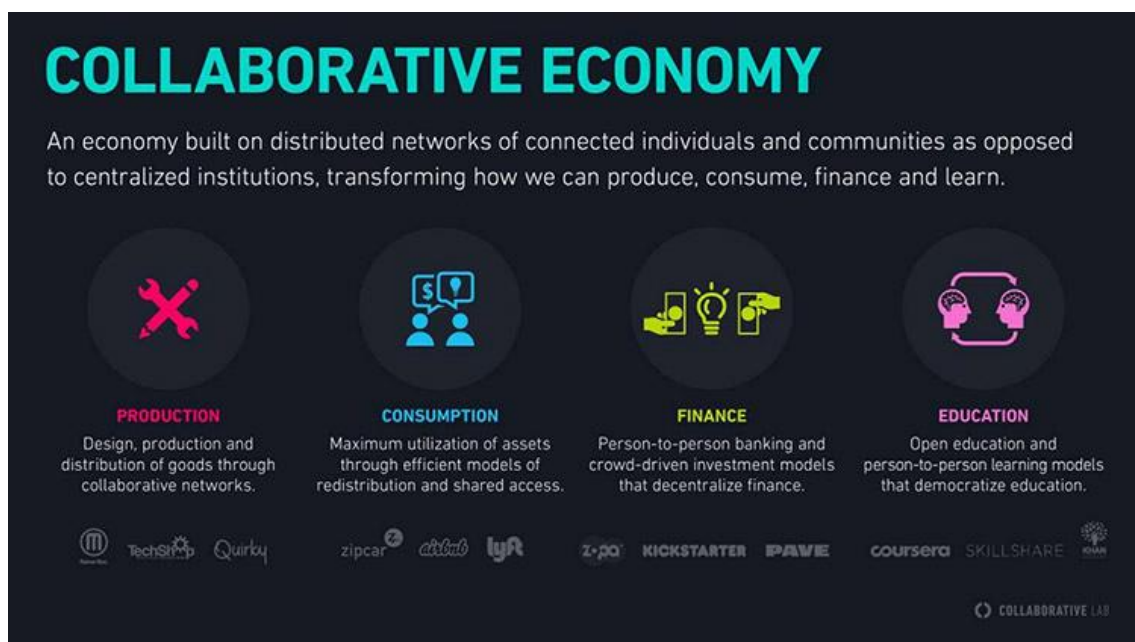


FIGURA 1. ECONOMÍA COLABORATIVA

La Producción Colaborativa o *peer production* es aquella que se da en cualquier ámbito en el que participan dos o más personas para desarrollar un proyecto en común. Acuñado por Yochai Benkler en 2006 [4], este sistema se basa en los principios del bien

común y del altruismo. Trasladado al ámbito digital, este tipo de producción abarca un abanico amplio de actores y logra coordinar los esfuerzos de toda una comunidad con el objetivo de obtener un fin común. Así, **la producción colaborativa forma parte de todos los usuarios de Internet**. Bien mediante las aportaciones a un proyecto, bien mediante la utilización de obras cuyo desarrollo ha sido posible gracias al esfuerzo de producción colaborativa. Es aquí donde se presenta un concepto estrechamente relacionado con la producción colaborativa: el Consumo Colaborativo, que podría definirse como el acceso a bienes y servicios a través de plataformas digitales. Bajo esta definición existen numerosas iniciativas: desde plataformas que buscan compartir de forma gratuita bienes escasos (Streetbank), hasta plataformas con un modelo de negocio más definido (Uber). Está presente también a través de los protocolos que manejan nuestras comunicaciones a través de Internet, en el software que utilizan algunos servidores, y en sitios reconocidos como YouTube, Flickr o Wikipedia.

Como ocurre con otros muchos conceptos relativos al ámbito (conocimiento abierto, finanzas compartidas), no sigue una estructura jerárquica tradicional. Si bien es cierto que, como norma general, existen una o más personas que coordinan los esfuerzos de la comunidad que desarrolla un proyecto, también lo es que **cualquiera que esté interesado en aportar puede hacerlo de diversas maneras**. No es un modelo de producción centralizado, ya que su valor radica precisamente en la diversidad de opiniones, puntos de vista y conocimientos que un colectivo aporta para beneficio del proyecto conjunto. Es un modelo de producción en el que toda colaboración es valorada por los actores que intervienen en el esfuerzo colectivo.

El auge de este modelo ha creado todo un modelo de negocios alrededor del mismo. El término Economía Colaborativa proviene de la expresión inglesa *Sharing Economy*, divulgado por Lisa Gansky [5] y Rachel Bootsman con Roo Rogers [6] en 2010. Un ejemplo de convergencia de estos conceptos es la **producción de software libre**, en la que expertos y entusiastas de la informática desarrollan conjuntamente herramientas y programas que pueden ser utilizados libremente por la colectividad.

No obstante, la producción colaborativa no se limita únicamente a la elaboración de software libre o de otros servicios profesionales (como pueden ser los educativos). Del mismo modo que el consumo colaborativo abarca iniciativas de cooperación desinteresada junto con modelos de negocio lucrativo. Aprovechando este modelo de producción, desarrolladores y empresas han creado un mercado de millones de dólares en la venta tanto de productos como de servicios (cabe aclarar que software libre no implica necesariamente gratuito).

La producción colaborativa también se encuentra en la creación de contenidos artísticos y culturales. El ejemplo clásico es Wikipedia [7]. Wikipedia es una enciclopedia en línea desarrollada gracias al esfuerzo de miles de colaboradores que han contribuido a crear, mejorar y corregir artículos en un gran número de idiomas. Aunque Wikipedia nace como resultado del esfuerzo colaborativo sin ánimo de lucro, las donaciones que

recibe la Fundación Wikimedia [8] la colocan como un modelo de negocio exitoso a partir de la producción colaborativa.



FIGURA 2. DONACIONES A WIKIMEDIA ENTRE EL AÑO 2013 Y 2014

Otro ejemplo es ccMixter [9]. Este sitio dispone de un gran número de obras musicales que permiten la mezcla de las mismas con el objetivo de producir nuevas obras. Algunas permiten también su utilización comercial, por lo que se puede considerar que forma parte de esta nueva economía.

Estos nuevos conceptos también han servido como motor para impulsar un cambio en la forma en que entendemos los derechos de autor. Estos derechos no son manejados en el mundo digital de la misma forma que los entendemos en el mundo analógico. Como consecuencia, las diferentes comunidades de producción colaborativa han diseñado esquemas de licencias que **adecúan el sistema de derechos de autor a las necesidades de la producción colaborativa** [10]. Incluso han desarrollado modelos que garantizan que las obras producto de esfuerzos colaborativos permanezcan abiertas, previniendo su apropiación por parte de particulares que puedan pretender transformarlas en un modelo de obra propietario. Es decir, el autor tiene libertad de decisión sobre su obra.

Es de prever que en el futuro coexistirán, más de lo que ya vemos ahora, modelos de trabajo basados en el lucro personal y empresarial con otros más altruistas, orientados a la cooperación y el desarrollo en todos los ámbitos de la economía.

## 2.2. COMUNIDADES COLABORATIVAS

Como explicamos en el apartado anterior, Wikipedia [11] es una de las plataformas colaborativas más populares. Wikipedia, junto con el resto de proyectos incluidos en Wikimedia Foundation, se desarrolla a un ritmo de 10 ediciones por segundo y, en concreto, Wikipedia en inglés tiene un promedio de 800 artículos por día [12]. El

estudio de Felipe Ortega, Jesús M. González-Barahona y Gregorio Robles [13] se centra en uno de los puntos que queremos tratar en nuestro proyecto: cómo de iguales (o desiguales) son las contribuciones de usuarios en una de las plataformas colaborativas más importantes, Wikipedia.

En dicho artículo, los autores encuentran una enorme desigualdad en todos los lenguajes de Wikipedia tratados en el estudio, en el que menos del 10% del total de autores son responsables de más del 90% del total de contribuciones. Este comportamiento se da en todos los lenguajes estudiados. Por ejemplo, la Wikipedia en japonés, que ocupa el 4º puesto en cuanto al número de artículos, y el 5º en relación a su número total de autores, es la comunidad que presenta un patrón de igualdad más alto respecto al número de contribuciones. Otras ediciones, como la alemana y sobretodo la inglesa, que cuenta con un número muy elevado de autores, muestran un nivel más alto de desigualdad. Esto refleja que el número de autores que crean artículos o el total de artículos modificados no son los factores que determinan el nivel de desigualdad en una comunidad. Otra conclusión que comparten los autores es la que indica que la evolución en el tiempo de esa desigualdad ha permanecido prácticamente estable a lo largo de los meses (para todos los lenguajes).

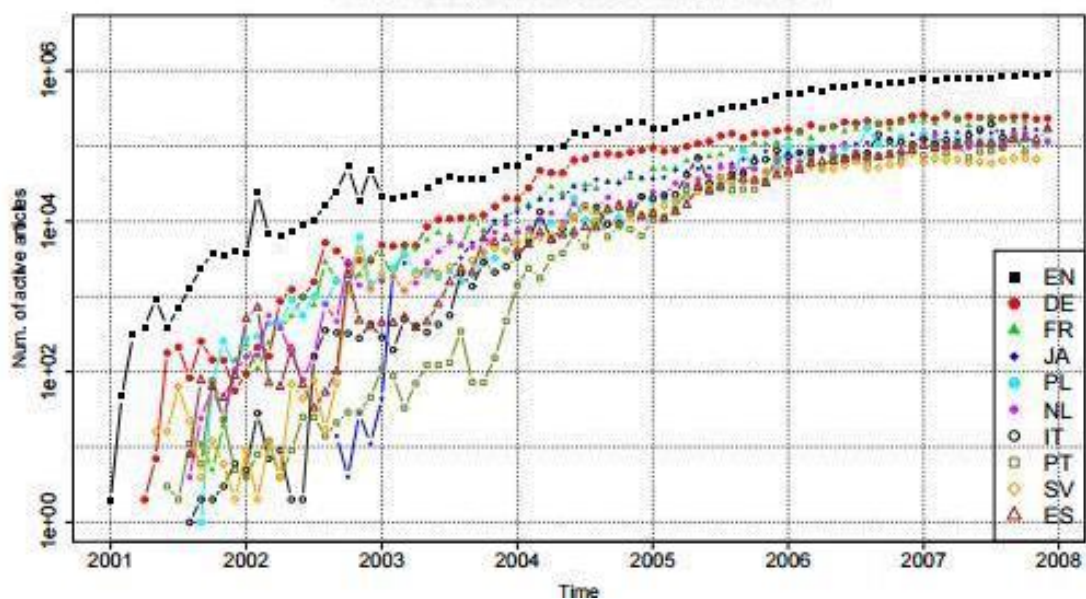


FIGURA 3. EVOLUCIÓN DEL NÚMERO DE ARTÍCULOS ACTIVOS EN FUNCIÓN DEL IDIOMA

No obstante, no es la única comunidad con una actividad elevada. StackOverflow recibe en torno a 32 millones de visitas mensuales a nivel global [14]. Esta comunidad, que proporciona un espacio para la consulta y resolución de cuestiones relacionadas sobretodo (aunque no exclusivamente) con la informática, dispone de un sistema de moderación basado en la reputación: todos los usuarios pueden alcanzar cierto poder de moderación en función de la reputación que tengan. Así cuantos más puntos tenga el usuario, más cosas puede hacer. De esta manera se divide el coste de moderar entre toda

la comunidad. Sin embargo, como ocurre con la mayoría de comunidades, el volumen de usuarios *core* es bajo comparado con la popularidad de la plataforma. Un estudio de 2013 [15] llegó a la conclusión de que el 77% de los usuarios sólo hacen una pregunta, el 65% responden únicamente a una, y **sólo el 8% responde a más de 5**.

### 2.2.1. WIKIA

Wikia [16] está catalogada como una fuente de información dinámica creada por fans, que abarca prácticamente todos los aspectos de la cultura popular: juegos de consola, programas de TV, festivales de música, grandes franquicias de películas, eventos deportivos, lanzamientos anticipados de libros, tendencias de moda y acontecimientos actuales a escala internacional. La base de fans de Wikia continúa alimentando un crecimiento consistente con más de 2 billones de visitas mensuales de páginas globales en más de 200 idiomas, y alojando más de 400,000 comunidades para fans [17].

Por tanto, podemos definir Wikia como un **conjunto de enciclopedias colaborativas online cuyo objetivo principal es cubrir la información y debate sobre temas de cultura popular** a un nivel de detalle mucho mayor que el que se puede encontrar en artículos de Wikipedia debido a sus políticas de uso. Así, cada enciclopedia constituye una wiki: un sitio web cuyas páginas pueden ser editadas directamente desde el navegador, donde los usuarios crean, modifican o eliminan contenidos que comparten. Wikia utiliza el sistema MediaWiki, software de código abierto escrito en PHP (en el que se basan muchos proyectos de la Fundación Wikimedia) para alojar dichas comunidades o wikis.

Además, el estilo de escritura está dirigido sobre todo a usuarios familiarizados con vocabulario y terminología específica, en vez de al público en general. Por ejemplo, un personaje secundario del videojuego *The Last Of Us* puede tener su propio artículo en la wiki de Wikia *The Last Of Us*, pero no puede considerarse lo suficientemente importante como para dedicarle una página de Wikipedia.

Dentro de toda la información contenida en las wikis es de especial importancia la obtención de información relacionada con los usuarios y su actividad: qué se puede obtener sobre ellos, cuánto son activos, qué nivel de actividad tienen, si es continuada en el tiempo... Los aportes hacen referencia a las **ediciones de cualquier tipo que un usuario hace en una wiki**. Así, si un usuario escribe dos comentarios en un artículo de una wiki constituye un solo aporte a esa wiki, pero dos contribuciones.

Wikia tiene varios tipos de usuarios, pero nosotros trabajaremos con los siguientes:

- **Administradores.** Son usuarios de confianza, generalmente elegidos por la comunidad. Tienen los mismos permisos que los burócratas y reversores, y además pueden borrar y recuperar páginas, bloquean páginas para evitar que se editen sin permisos de administrador, pueden bloquear usuarios, otorgan y

- revocan derechos de moderador de chat y pueden editar el diseño y tema de una wikia.
- **Burócratas.** Los burócratas se encuentran un nivel más alto que los administradores, pero no tienen ninguno de los privilegios de un administrador. Solo tienen la habilidad de nombrar a otros usuarios como burócratas, y pueden conceder y revocar permisos de administrador y de reversor a los usuarios de su wikia.
  - **Reversores.** Los usuarios con permiso de reversor son capaces de deshacer malas ediciones con un solo clic, utilizando el enlace de reversión que aparece en la diferencia de páginas y en la lista de contribuciones del usuario.

El resto de categorías de usuarios no existen en todas las comunidades y registran poca actividad.

Además, los usuarios se pueden clasificar en función de su actividad dentro la comunidad. Esto es conocido como la Teoría 90-9-1 o Desigualdad Participativa [18], que estipula que todo sitio que precisa de la colaboración de una comunidad para su funcionamiento sufre una desigualdad en dicha participación, la cual se representa por, aproximadamente, las siguientes proporciones:

- El 90% de los usuarios son usuarios/consumidores. Se dedican a observar, pero no aportan contenido alguno. No obstante, participan dando difusión a artículos, compartiendo en redes sociales, etc.
- El 9% de los usuarios contribuye ocasionalmente, representando el 10% del contenido de la plataforma.
- El 1% de los usuarios se atribuye más del 90% de las participaciones y de otras actividades del sistema. Este 1% recibe el nombre de usuarios core (*heavy contributors*), ya que su actividad supera por mucho a la del resto de integrantes.

No obstante, en Wikia un usuario se considera activo cuando **ha participado al menos una vez en los últimos 30 días**.

Los usuarios proporcionan cierta información cuando se registran en Wikia. Parte de ella (tipo de usuario, actividad) queda registrada y es fácilmente accesible (ver sección 6.1). Pero algunos de los datos que se solicitan a los usuarios son opcionales (edad, procedencia, fecha de nacimiento, ocupación y género), por lo que la mayoría no los completan y, aun pudiendo trabajar con ellos, no representan la realidad por ser una muestra tan pequeña.

En línea con la participación, las diversas wikis disponen de logros que pueden ser habilitados por los administradores de cada comunidad. Estas insignias se otorgan a los usuarios y fomentan la colaboración, premian a sus miembros y promueven el crecimiento de la comunidad. Analizar de qué forma se otorgan logros así como quién los recibe, entre otros, es una forma de establecer conexiones entre usuarios y entre wikis.

Con el objetivo de encontrar más información útil y explotarla en nuestra aplicación, consultamos también los rankings de Alexa para Wikia [19]. Por ejemplo, el perfil de los visitantes o cuáles son los países que más consultas realizan.

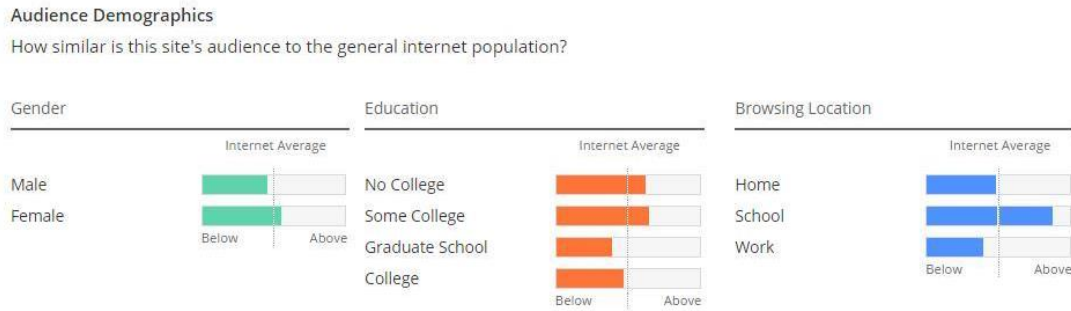


FIGURA 4. PERFIL DE USUARIOS QUE MÁS VISITAS REALIZAN A WIKIA

Elegimos centrarnos en Wikia y, en concreto, Wikia en español, por los siguientes motivos:

- Era necesario trabajar con un subconjunto de comunidades por no poder escalar a toda Wikia
- Filtrar por idioma permite maximizar diversidad
- El subconjunto de todas las wikis contiene una gran cantidad de información con la que sería útil trabajar.
- La interpretación y el análisis de esos datos, así como su extracción, es similar entre unas páginas y otras.

Además, la mayor parte de los textos de Wikia están disponibles para su descarga bajo una licencia Creative Commons, facilitando el acceso a la información. Esta característica nos lleva al punto siguiente.

## 2.3. ACCESO A DATOS

Los motivos que se hallan tras el avance de las herramientas de visualización de datos, como comentábamos en el Capítulo 1, parten de un mismo punto en común: la necesidad de simplificar la presentación de datos e información relevante. Éste ya es un motivo de envergadura, que por sí mismo justifica el buen momento que atraviesan este tipo de instrumentos. Es por ello que ya hay un gran número de herramientas de visualización de datos disponibles.

El primer paso a la hora de llevar a cabo un trabajo como el que se persigue en este proyecto es disponer de datos. Aquí entra el concepto de Open Data.

El Open Data es el movimiento digital que persigue un **acceso libre a la información** y datos con los que trabajan gobiernos e instituciones, sin restricciones de copyright, patentes u otros mecanismos de control, en formatos digitales y estandarizados que permitan su reutilización [20]. La libre disposición de información aporta transparencia de la gestión de organismos públicos y posibilita la intervención ciudadana en la toma



de decisiones de proyectos o iniciativas. Un ejemplo es el de *¿Dónde van mis impuestos?* [21], un servicio que explica los gastos anuales de la Administración Central del Estado y la Seguridad Social, tal como aparecen recogidos en los presupuestos generales.

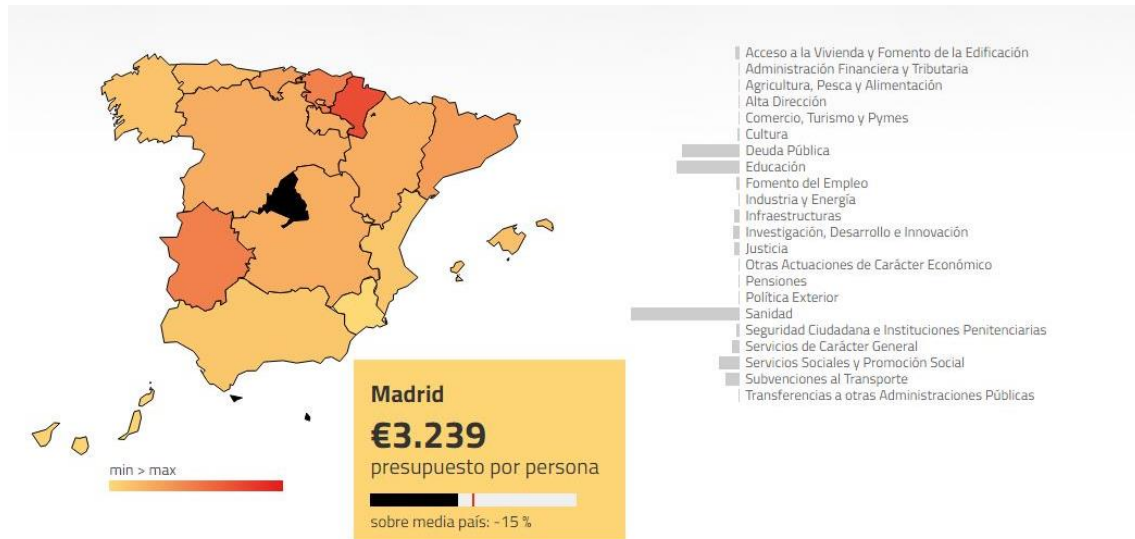


FIGURA 5. DISTRIBUCIÓN DEL PRESUPUESTO EN LA COMUNIDAD DE MADRID PARA EL AÑO 2015

El objetivo es que ciudadanos y empresas puedan reaprovechar estos datos para generar valor económico. Podrán construir sobre ellos una nueva idea que genere nuevos datos, conocimientos o incluso la creación de nuevos servicios que reporten beneficios económicos o sociales.

El World Wide Web Consortium, la comunidad internacional que desarrolla estándares para el crecimiento de la web, recomienda que **los datos que se liberen sean útiles y estén en formatos reutilizables** [22]. Cuanto mejor estructurados y enriquecidos estén los datos, más fácil será reutilizarlos y construir aplicaciones que puedan tratarlos automáticamente. También es importante que los datos sean actuales y detallados, sin procesar y manteniendo el mayor nivel de detalle posible. Estos datos en bruto se organizan en datasets, o catálogos de datos, con el objetivo de facilitar su localización y ser más fácilmente indexados. Por ello, se utilizan campos que definen el grupo de datos como la descripción, la frecuencia de actualización, el formato o la licencia de uso entre otros. Los formatos [23] más utilizados son:

- RDF: es un modelo universal para representar datos en la Web, que permite intercambiar y enlazar información y recursos a través de diferentes aplicaciones, lo que facilita la reutilización y enriquecimiento de los recursos en la Web.
- CSV: es un tipo de documento plano cuyos valores organizados en columnas (separados por comas) y filas (saltos de línea).



- XML: es un lenguaje de marcado que permite la interpretación de datos para diferentes lenguajes. Es el estándar para el intercambio de información estructurada entre diferentes plataformas (bases de datos, hojas de cálculo)
- ODATA: es un protocolo abierto que permite realizar consultas y selecciones sobre los conjuntos de datos ofrecidos, así como descargar los resultados de esas operaciones.
- JSON: basado en la notación literal de objetos de JavaScript. Su sintaxis es sencilla, por lo que facilita el tratamiento en los navegadores.

Actualmente existen muchas organizaciones que proporcionan datasets de distintas categorías, aportando visibilidad a la información y permitiendo su reutilización.

## Open Collaboration Data Factories

Entre ellos encontramos la comunidad Open Collaboration Data Factories [24], que proporciona datasets de comunidades online con un estándar de datos. Este proyecto tiene como objetivo facilitar el acceso a la información relacionada con medios sociales y comunidades online, agregando datasets públicos relativos a estos temas. El uso de un estándar de metadatos es un proyecto en curso que pretende identificar y catalogar fuentes de datos abiertos.

Aunque acaban de cambiar al estándar 3.0 y en este aún no hay muchos ficheros, dentro del 2.0 hay un amplio abanico disponible. Accediendo a su repositorio encontramos los enlaces desde donde se pueden descargar los archivos. La mayoría de ellos están en texto plano, aunque hay algunos en formato JSON o SQL.

## Open Referral

La salud y los servicios sociales son recursos esenciales que ayudan a las personas a satisfacer sus necesidades y mejorar su calidad de vida. El estándar Human Services Data Specification (HSDS) está siendo desarrollado por la iniciativa Open Referral [25] con el objetivo de que la información sobre tales servicios sea más fácil de mantener, compartir, encontrar y utilizar. *Code for America* dispone de una organización dentro de Github donde está definida la especificación de dicho estándar [26].

## DERP

The Digital Ecologies Research Partnership [27], DERP, es una iniciativa conjunta de una alianza de sitios web que promueve un acceso a la investigación académica abierta y accesible al público. Se podría decir, por tanto, que DERP se centra en proporcionar datos públicos a investigadores académicos. Esta plataforma apoya la investigación que respete la privacidad del usuario, y toda investigación que reciba su respaldo se dará a conocer de manera abierta. Algunas organizaciones, como Reddit o Imgur, colaboran proporcionando datos y apoyando la investigación académica.

## 2.4. HERRAMIENTAS DE VISUALIZACIÓN Y ANÁLISIS DE INFORMACIÓN

Cuando se dispone de tanta información como en la actualidad y queremos tratarla no podemos limitarnos a obtener datos sin llegar a una correcta comprensión de los mismos. Es en ese punto donde entran en juego las herramientas de tratamiento de datos. En concreto, nos centramos en las herramientas de visualización y análisis.

La visualización [28] es una disciplina que utiliza el poder de comunicación de las imágenes para explicar de manera comprensible las relaciones de significado y dependencia que se pueden encontrar entre las grandes masas de información.

Cuando trabajamos en torno a la visualización de datos en numerosas ocasiones queremos ser creativos y huir de los “aburridos” gráficos de barras. Sin embargo, el hecho de que dispongamos de formas de visualización más atractivas no quiere decir que sean las más efectivas siempre. Los gráficos de burbujas facilitan más puntos de datos en menos espacio, los gráficos de anillos indican claramente las relaciones parte-todo, y los diagramas de árbol muestran fácilmente las categorías jerárquicas. Pero ninguno de ellos es más efectivo que un simple gráfico de barras para una comparación entre valores. Debemos pensar siempre en **facilitar la comprensión sin distorsión de los datos** facilitados. Los gráficos de barras son una de las mejores herramientas disponibles para facilitar las comparaciones visuales, aprovechando nuestra capacidad innata para comparar con precisión longitudes que están una al lado de la otra.

A continuación se describen algunas de las herramientas de visualización clasificadas de acuerdo a sus características y a la similitud respecto al propósito de este proyecto.

### 2.4.1. HERRAMIENTAS DE ANÁLISIS ESTADÍSTICO

Igual de importante que la visualización es poder tratarla: reconocer patrones, establecer relaciones entre los datos y predecir un determinado comportamiento. Esto puede llevar cierto tiempo, pero resulta más sencillo si se sabe, por ejemplo, qué variables comprobar, y se dispone de las herramientas adecuadas.

#### WAM de Wikia

Una de los servicios que ofrece Wikia es el WAM. Este monitor de actividad utiliza un algoritmo que constituye una forma de evaluación para administradores, fundadores y editores que quieran entender mejor como se sitúan sus wikis en cuanto a actividad con respecto a otras comunidades.

Cualquiera puede consultar este ranking a través del enlace principal de Wikia y en las páginas principales de videojuegos o entretenimiento. WAM calcula diariamente las puntuaciones para las 5.000 comunidades de Wikia más visitadas y hace una

clasificación general y vertical con puntuaciones del 0 al 100. La clasificación del WAM es sensible a los eventos del mundo real y cambia frecuentemente, lo que permite estar actualizados en relación a los últimos cambios en las clasificaciones. Las wikis analizadas se dividen en categorías (TV, Libros, Comics,...) y aportan información sobre sus administradores.

Rank	WAM Score	Peak Rank	Wikia URL	Vertical	Vertical Rank	Admins
15	● 99.36	1	<a href="http://es.drama.wikia.com">es.drama.wikia.com</a>	TV	3	
27	● 99.12	1	<a href="http://es.doblaje.wikia.com">es.doblaje.wikia.com</a>	TV	5	
40	↓ 98.84	6	<a href="http://es.pokemon.wikia.com">es.pokemon.wikia.com</a>	Games	21	
51	↓ 98.42	1	<a href="http://hieloyfuego.wikia.com">hieloyfuego.wikia.com</a>	Books	2	
94	↓ 97.29	3	<a href="http://es.naruto.wikia.com">es.naruto.wikia.com</a>	Comics	12	
96	● 97.28	15	<a href="http://es.terraria.wikia.com">es.terraria.wikia.com</a>	Games	48	
103	↓ 97.16	1	<a href="http://es.steven-universe.wikia.com">es.steven-universe.wikia.com</a>	TV	26	
120	↓ 96.68	33	<a href="http://es.yugioh.wikia.com">es.yugioh.wikia.com</a>	Games	57	
127	↓ 96.49	2	<a href="http://es.dragonball.wikia.com">es.dragonball.wikia.com</a>	Comics	15	
157	↓ 95.73	7	<a href="http://inciclopedia.wikia.com">inciclopedia.wikia.com</a>	Lifestyle	14	
170	↓ 95.49	4	<a href="http://es.onepiece.wikia.com">es.onepiece.wikia.com</a>	Comics	18	
183	↓ 95.11	3	<a href="http://es.horadeaventura.wikia.com">es.horadeaventura.wikia.com</a>	TV	49	
192	↑ 94.93	11	<a href="http://es.youtube.wikia.com">es.youtube.wikia.com</a>	TV	51	

FIGURA 6. CLASIFICACIÓN WAM PARA LENGUAJE ESPAÑOL A 2 DE JUNIO DE 2016

Es decir, esta clasificación se basa en el tráfico, el compromiso y el crecimiento de cada wiki. Para subir posiciones en ese ranking lo más importante es **fomentar la colaboración**. Por ejemplo, las wikis con un número elevado de editores activos funcionan mejor en la clasificación.

Parte de la información registrada en el WAM nos sirve de guía para desarrollar nuestra aplicación, teniendo en cuenta, por ejemplo, la información sobre los administradores de cada comunidad.

## Quantcast

Quantcast [29] es una empresa fundada en 2006, especializada en la medición de audiencias y la publicidad en tiempo real. La compañía ofrece el acceso público a los datos de tráfico y demográficos para millones de sitios web. Sus centros de datos procesan más de 800.000 transacciones por segundo [30] y la compañía indica que produce una precisa medición de audiencia a más de 100 millones de destinos web. A partir de 2013 pasó a formar parte de las cinco principales organizaciones de procesamiento de datos a escala mundial.

## Alexa Ranking

Filial de la compañía Amazon, este servicio provee información acerca de la cantidad de visitas que recibe un sitio web y los clasifica en un ranking [31]. Además, muestra gráficas donde se puede apreciar el crecimiento/decrecimiento de las visitas a una página web. La interfaz es limpia y clara, pero el software solo permite realizar consultas gratuitas de un cierto nivel.

## Bitergia Cauldron

Bitergia es una compañía apoyada por el Vivero de Empresas del Parque Científico de la Universidad Carlos III de Madrid [32] y que se define como una empresa de análisis de datos de comunidades de desarrollo de software libre. Como servicio básico se proporciona un análisis completo y visualización de proyectos de desarrollo de software en una interfaz de usuario.

Cauldron [33] es un servicio PoC diseñado por Bitergia para el análisis de la actividad (*git commits*, *topics* y *pull requests*) en un propietario GitHub, bien usuario u organización. Actualmente, Cauldron produce un dashboard para visualizar los principales parámetros de los 10 repositorios públicos con más actividad reciente para el propietario especificado utilizando el toolkit GrimoireLab.

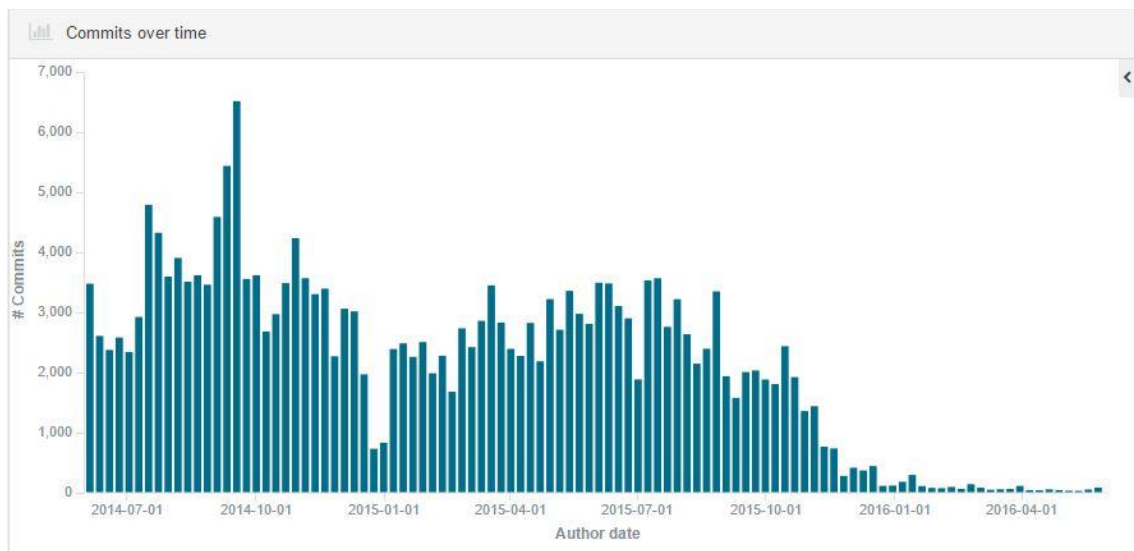


FIGURA 7. *COMMITTS* A LO LARGO DEL TIEMPO PARA ANDROID DE CAULDRON

Además, otros componentes fundamentales para el desarrollo de este dashboard son Python (casi todo el código de análisis utiliza este lenguaje), Elasticsearch (almacenamiento) y Kibana (concretamente, Kibitter, fork de Kibana)

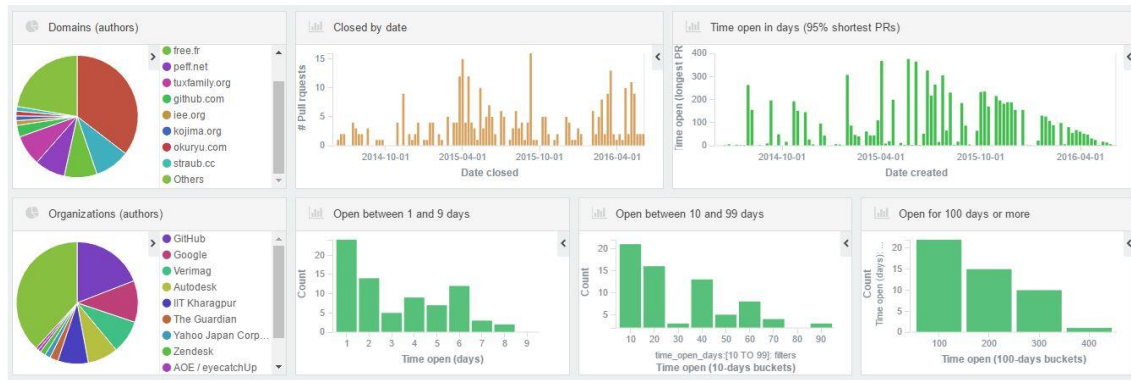


FIGURA 8. PANEL DE *PULL REQUESTS* PARA GIT DE CAULDRON

Los dashboards de Cauldron agrupan esa información en una serie de paneles: Overview, Issues, Git y Pull Requests:

- Overview muestra una descripción general de los repositorios analizados.
- El panel Git ofrece información sobre commits: quién es el autor y cuándo se hizo el commit, y quién lo incluyó en el repositorio y cuándo. Utilizan la información de la cuenta de correo para identificar el dominio. También muestran información horaria respecto a la zona en la que se hizo la contribución.
- El panel de Issues muestra información relativa a los problemas detectados, excluyendo los pull requests. Cada *issue* es creado por una persona, y puede ser asignado y/o resuelto por otras. Para cada una de estas acciones, la API de GitHub proporciona una fecha, de manera que es posible controlar la solución de esos *issues* a lo largo del tiempo.
- El panel de Pull requests informa sobre este tipo especial de *issue*.

Cuando el usuario interactúa con esos paneles, la información disponible en el conjunto del panel (o en todo el dashboard) se recalcula y reconfigura, filtrando parte de los datos. Además, todo se puede guardar y compartir (manteniendo los filtros).

Tanto Quantcast como los rankings de Alexa o WAM nos proporcionan ideas a la hora de obtener información de Wikia (como cuáles son las comunidades más populares), pero no nos sirven de referencia como herramienta, ya que uno de nuestros objetivos es dar al usuario la opción de “jugar” con la información y no solo consultarla.

Nuestra aplicación pretende, por tanto, seguir los pasos de Cauldron, mostrando la información más importante obtenida de las comunidades de Wikia en español y permitiendo al usuario interactuar con ella.

## 2.4.2. SERVICIOS DE VISUALIZACIÓN

Internet podría considerarse una autopista de información en la que la web facilita el acceso a los datos disponibles. Los navegadores son, por tanto, los programas que actúan como portal a la web, facilitando el acceso a la información y constituyendo una herramienta de visualización muy potente.

### 2.4.2.1. Aplicaciones de visualización

Esta sección incluye herramientas que ofrecen diversas opciones de visualización. Aunque algunas siguen apostando por las tablas y gráficos clásicos, muchas otras abogan por ofrecer nuevas prestaciones, tales como diagramas de árbol y nubes de palabras.

#### Google Fusion Tables

Google, la empresa principal subsidiaria de la multinacional Alphabet Inc., también ha desarrollado su propia herramienta de visualización de datos [34]. Solo requiere una cuenta de Google, y permite compartir los datos de forma abierta y construir gráficos personalizados de visualización. Gestiona grandes colecciones de datos que deben estar normalizados y guardados en un archivo Excel, .ods, .csv o .kml, y proporciona un medio para visualizar, no sólo gráficos de barras o circulares, sino también mapas geográficos basados en Google Maps.

Sacarle el máximo partido a esta herramienta requiere cierta práctica y, aunque los menús de ayuda principales están dirigidos a personas con cierto control de Excel y Google Maps, los submenús redirigen eventualmente a las secciones que proporcionan la ayuda adecuada.

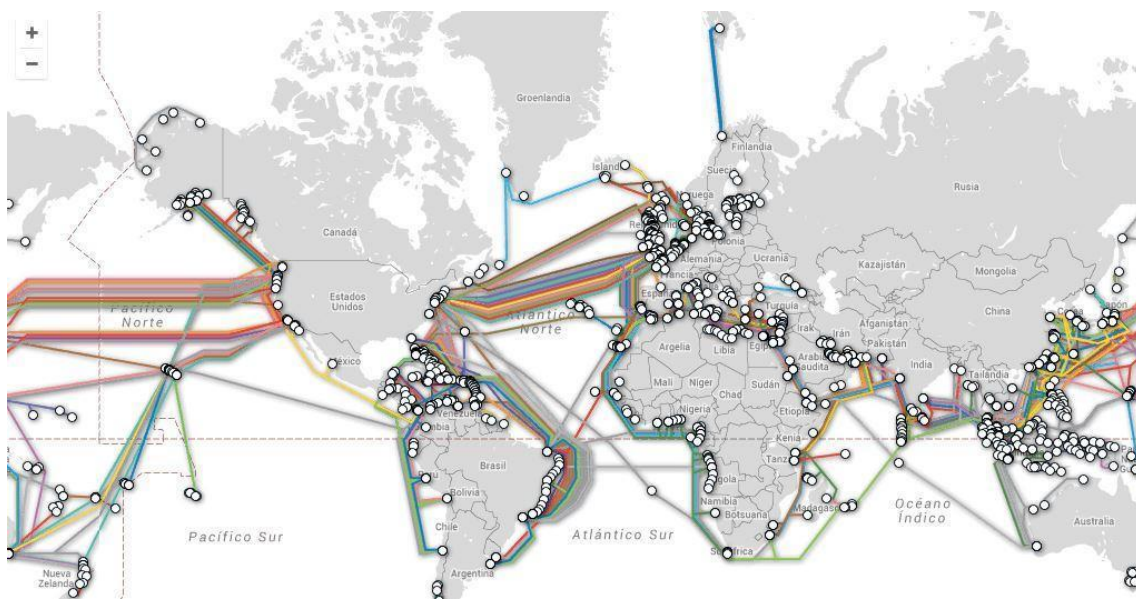


FIGURA 9. GOOGLE FUSION TABLES: THE SUBMARINE CABLE MAP



## CartoDB

La finalidad de CartoDB [35] es facilitar la creación de aplicaciones de geolocalización y la creación de mapas. Permite diseñar y desarrollar mapas en tiempo real que funcionan en todas las plataformas web y móviles. Dispone de una base de datos geoespacial en la nube, que funciona con los servicios web de Amazon, permitiendo la escalabilidad, la flexibilidad y la elasticidad de sus servicios. Es un proyecto Open Source que también se ofrece como un servicio bajo demanda.

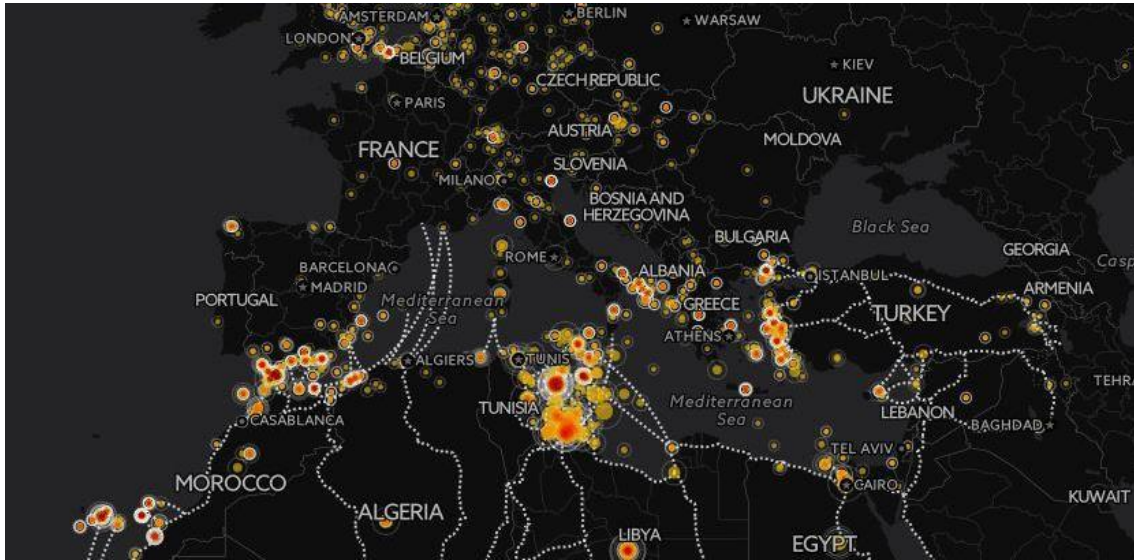


FIGURA 10. CARTODB: IMPACTO DEL FLUJO DE INMIGRANTES EN POBLACIÓN Y MUERTES (2000-2015)

Está orientado a desarrolladores sin experiencia en sistemas de información geoespacial, con una interfaz muy amigable. Dispone de mucha documentación, tanto para usuarios nóveles como expertos. Sin embargo, si es conveniente tener conocimientos de SQL (PostgreSQL). Entre sus clientes se cuentan algunas instituciones tales como la ONU, Google, la NASA y la Universidad de Oxford.

## Tableau Software

Sin duda, Tableau [36] es una de las herramientas líder actualmente, tanto por la facilidad de uso como por sus funcionalidades. Su uso es muy intuitivo, y permite la creación de **visualizaciones de alto nivel y creación de informes** con tan sólo arrastrar los datos, mostrando los cambios en tiempo real. Es, además, rápida y permite manipular distintas representaciones de datos en un mismo dashboard. Ofrece también la posibilidad de añadir información extra, mediante de la inserción de documentos o páginas web, que mejora la comprensión de la información.

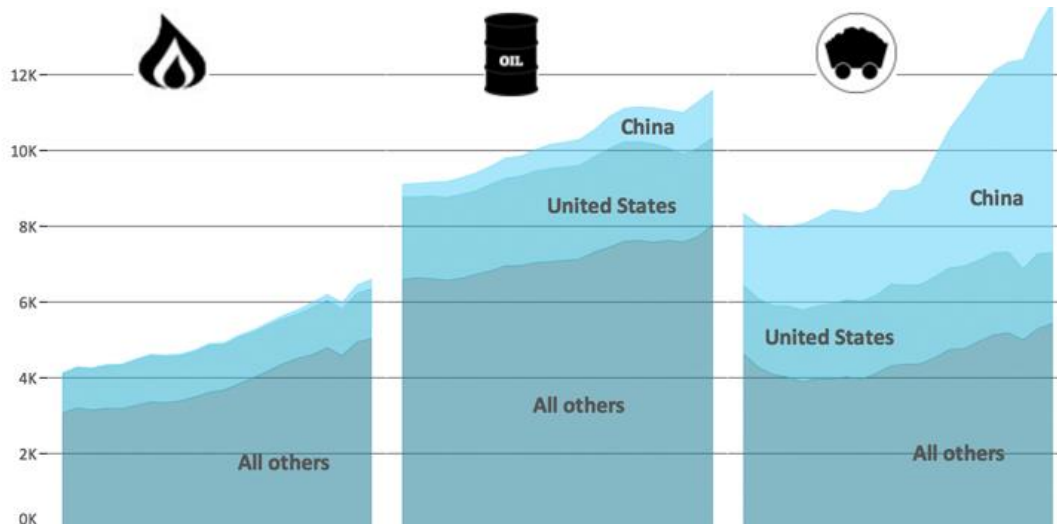


FIGURA 11. TABLEAU: EMISIONES MUNDIALES DE CO2 POR GAS Y PETRÓLEO

Si los datos con los que se desea trabajar están en Excel o en CSV, el usuario puede subir el fichero de manera sencilla a Tableau. Sin embargo, si desea conectarse a una base de datos, un desarrollador con conocimientos en SQL tendrá que crear una consulta para *tirar* del dataset. Así, a pesar de que Tableau da a los usuarios amplia libertad una vez creado el dataset, es necesario conocer SQL para crearlo inicialmente.

#### 2.4.2.2. Librerías y frameworks

Podríamos definir el término librería como un conjunto de código estandarizado y organizado que contiene funciones que facilitan la programación. Así, un *framework* puede verse como una gran librería o conjunto de librerías donde, además de facilitar funciones para su uso, se suele disponer de una sintaxis y una forma de organización del código específicas.

Para los desarrolladores web, sobre todo tipo *front-end*, las librerías se han convertido en un elemento fundamental para implementar características dinámicas en diversos proyectos. Todos los navegadores, además, llevan integrados sus propios motores JavaScript con el objetivo de aprovechar al máximo sus características y ofrecer al usuario un mayor rendimiento en aplicaciones web.

JavaScript dispone de un gran número de librerías y frameworks cuyas funcionalidades se centran en la visualización de gráficos. Estas librerías aportan mayor flexibilidad que las aplicaciones de escritorio (más restrictivas en cuanto a opciones), pero también requieren mucho más conocimiento por parte del usuario.



## D3.js

D3, *Data Driven Documents* [37], es una herramienta capaz de ofrecer visualizaciones interactivas online muy avanzadas con complejos conjuntos de datos. Se trata de una librería de JavaScript que nos da la posibilidad de crear unos diagramas completos y gráficos a partir de una amplia variedad de fuentes de datos. Básicamente, la librería permite manipular documentos basados en datos usando estándares abiertos de la web; y los navegadores pueden crear visualizaciones complejas sin depender de un software propietario. Sus desarrollos son abiertos y utiliza HTML, SVG y CSS para crear las visualizaciones. HTML y CSS son, dentro de las tecnologías que utiliza D3, las más sencillas de aprender. SVG es algo más complejo, pero esta librería dispone de un gran número de métodos de ayuda que facilitan la creación de elementos SVG. Y por último JavaScript, el lenguaje en el que se basa la librería, que tiene curva de aprendizaje.

Es decir, D3 es la herramienta perfecta para crear gráficos complejos y personalizados, pero el proceso de aprendizaje es más lento.

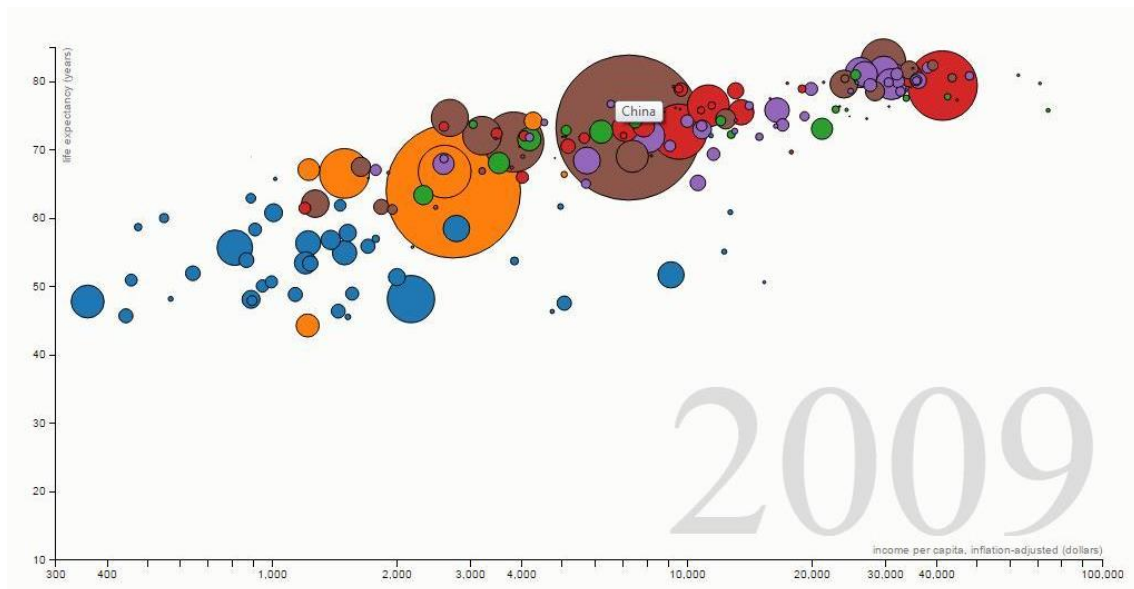


FIGURA 12. D3.JS: FLUCTUACIÓN DINÁMICA DE LOS INGRESOS PER CÁPITA (X), LA ESPERANZA DE VIDA (Y) Y POBLACIÓN (RADIO) DE 180 NACIONES EN LOS ÚLTIMOS 209 AÑOS

## Google Chart Tools

Google Chart Tools [38] es una herramienta de Google Developers que permite la creación de gráficos en forma de imágenes PNG. Su funcionamiento se basa en peticiones http a una determinada URL. Es de uso gratuito pero con ciertas limitaciones. Dispone de una gran variedad de tipos de gráficos, que vienen dados como clases de JavaScript. Una de las ventajas que tienen este sistema de generación de gráficos es que no se necesita instalar ningún componente en nuestro entorno o servidor.

Google Chart no ofrece unos gráficos tan variados como los que podemos crear con D3.js. Sin embargo, el proceso es mucho más sencillo: eliges el tipo de gráfico, introduces correctamente los datos y personalizas el estilo.

## Chart.js

Chart.js [39] es una librería JavaScript de software libre muy ligera (apenas 11kb). Incluye los tipos de gráficos líneas, barras, radar, polar y donut entre otros, cada uno con su propio módulo, permitiendo cargar sólo los que requiera un proyecto específico. Utiliza *canvas* HTML5 para representar los gráficos, todos interactivos.

La documentación que incluye Chart.js es clara y está bien organizada, y proporciona información detallada para el uso de cada una de sus características, aunque es necesario tener conocimientos de JavaScript y Ajax para realizar peticiones. Los gráficos son *responsive*, es decir, que se adaptan al tamaño disponible. Sin embargo, su principal desventaja es que solo ofrece 6 tipos de gráficos, que limita las opciones de visualización.

## Echarts.js

En Noviembre de 2014, la empresa china Baidu dio a conocer una nueva generación de librerías de gráficos [40]. Se llama ECharts (Enterprise Charts) y es una completa librería que ofrece una manera sencilla de añadir gráficos interactivos.

Está escrita enteramente en JavaScript y basada en ZRender. Con la funcionalidad Magic Switch es posible cambiar de un tipo de gráfico a otro similar, aumentando su versatilidad. El usuario puede, a través de un botón, modificar la apariencia de sus gráficas para adaptarla a las necesidades del momento.

Al igual que las anteriores librerías mencionadas, es necesario tener conocimientos de JavaScript y HTML para explotar al máximo la librería, aunque también dispone de una completa documentación y ejemplos.

Tras el análisis de estas y otras librerías de gráficos, nos decantamos por Echarts.js, por ser la que más funcionalidades ofrecía. Esta librería dispone de una API detallada, donde se describen todos los componentes que maneja y detalla cómo modificar los parámetros para obtener los resultados que se persiguen.

### 2.4.3. HERRAMIENTAS PARA EL ANÁLISIS DE REDES

Los datos generados en las redes sociales siempre están disponibles y la complejidad reside precisamente en saber cómo interpretarlos. Debido a esto, en los últimos años el análisis de datos se ha hecho imprescindible para generar comprensión de las dinámicas de las redes. Los siguientes son ejemplos de herramientas que facilitan la visualización de nodos entre redes.

#### NodeXL

Aplicación de escritorio que permite el análisis y representación de grafos que trabaja con Excel [41]. Muestra gráficos de red a partir de una lista dada de conexiones, ayudando en el análisis y descubrimiento de patrones y relaciones en los datos. Dispone de conexiones directas a las redes sociales, optimizada para analizar medios sociales en línea; incluyendo conexiones incorporadas para hacer *query* en las API de Twitter, Flickr y YouTube.

#### Gephi

Gephi [42] es una plataforma de exploración y visualización interactiva para todo tipo de redes y grafos complejos, dinámicos y jerárquicos. Esta herramienta es open source multiplataforma bajo la licencia GNU GPL 3 que permite procesar grandes volúmenes de datos y el manejo de gráficos en tiempo real. Su objetivo principal es ayudar a los analistas de datos a hacer hipótesis, aislar estructuras de datos o descubrir patrones de forma intuitiva. Su manipulación es bastante sencilla para el usuario, aunque es necesario conocer el código de la herramienta a la hora de utilizarla y explotarla al máximo. Los resultados obtenidos en los análisis pueden ser exportados en distintos formatos (PDF, SVG), que permite que esa información pueda ser reutilizada por otro software posteriormente.



FIGURA 13. GEPHI: TRAYECTORIA DE LOS REFUGIADOS SEGÚN EL INFORME DE UNITED NATIONS HIGH COMMISSIONER FOR REFUGEES

## 2.5. CONCLUSIONES

Como conclusión general, ante la gran cantidad y diversidad de herramientas de visualización existentes en el mercado, se puede inferir que estamos ante un periodo de gran proliferación de datos en bruto, en el que existe un creciente interés por encontrar la manera más adecuada de presentar esta información de manera atractiva, clara, concisa y comprensible para el usuario final.

Como última reflexión mencionar de nuevo que la plataforma que se sitúa más en línea con nuestro objetivo podría entenderse como una síntesis de Cauldron de Bitergia y parte del análisis que proporcionan Alexa o Quantcast mediante una aplicación web que hace uso de una librería de JavaScript para aportar dinamismo a los gráficos. Otras herramientas, como pueden ser de tipo Data Mining y de depuración, si bien muy útiles en el ámbito del procesamiento de los datos, así como las soluciones para el análisis de redes, se alejan del objetivo que se persigue en este proyecto.



# CAPÍTULO 3

## METODOLOGÍA DEL PROYECTO

Este capítulo se centra en el plan de trabajo seguido por el equipo durante el desarrollo de todo el proyecto, así como en su relación con el software libre.

### 3.1. PLAN DE TRABAJO

El desarrollo de nuestra aplicación se ha realizado a lo largo del curso por un equipo compuesto por tres personas bajo la tutela de dos directores de proyecto, acordado al terminar los exámenes de Junio del curso 2014/2015. Se mantuvieron reuniones con los tutores del trabajo para ir haciendo un seguimiento de los avances que habíamos conseguido, y para que nos aconsejaran cual era el siguiente paso a seguir. Cada vez que se concretaba algún avance, se informaba a todo el equipo mediante correo electrónico, sistema de comunicación que se empleaba también para agendar las reuniones. Además, había reuniones internas para discutir la línea de avance del proyecto y resolver los problemas o dudas que iban apareciendo.

El nivel de producción ha sufrido algunos altibajos y no ha sido constante durante todo el año:

- El primer trimestre (Octubre-Diciembre) del curso el trabajo nos centramos en la investigación de las tecnologías que íbamos a utilizar, como librerías de gráficos, y en desarrollar los primeros bocetos de la aplicación web, centrándonos en la metodología del Diseño Guiado por Objetivos. A finales de Noviembre iniciamos el desarrollo de la herramienta, mostrando únicamente algunos de los posibles gráficos y con datos ficticios.
- El segundo trimestre (Enero-Marzo), coincidiendo con exámenes, la dedicación al proyecto bajó, de manera que los principales aportes se centraron en ampliar la memoria con los avances conseguidos hasta la fecha. En el mes de Marzo la dedicación volvió a aumentar, y la herramienta ya mostraba datos reales hacia mediados del mes, con nuevas funcionalidades.
- El último trimestre (Abril-Junio) impulsamos notablemente el avance del proyecto. Realizamos una serie de evaluaciones con usuarios utilizando la

primera versión de la aplicación y nos centramos en mejorar en línea con lo que nos habían sugerido dichas evaluaciones.

El paso fundamental con respecto a la organización del proyecto fue el de elegir si queríamos trabajar con una comunidad colaborativa o con varias. Dado que, como ya comentamos antes, no siguen un estándar a la hora de almacenar información, tuvimos que elegir entre, o bien trabajar a un nivel muy básico con varias plataformas, o profundizar más en la obtención de los datos de una única candidata. Escogimos centrarnos en Wikia (ver sección 2.2.1) e intentar cubrir un ciclo completo.

La organización del proyecto quedó, por tanto, estructurada de la siguiente manera:

1. Desarrollar una herramienta que, a partir de una wiki, muestre una determinada información básica.
2. Mejorar la aplicación para que muestre más de una wiki.
3. Mejorar la aplicación para que muestre mayor cantidad de datos e interactividad al usuario.
4. Conseguir comparar información entre distintas wikis.
5. Trabajar esa información, tratando de hacer la visualización y análisis reusable para otras plataformas (ganar en abstracción)

## 3.2. ENFOQUE DE SOFTWARE LIBRE

Dado que se trata de un proyecto de análisis de información de comunidades colaborativas, no tendría ningún sentido que tanto este documento como la aplicación web resultante no fueran libres. Por tanto, todo el código fuente mencionado en este documento es accesible a través de GitHub dentro de la organización del grupo de investigación GRASIA de la UCM:

<https://github.com/Grasia/ChartsUp>

La licencia elegida para nuestro software es una **licencia MIT**, una de las numerosas licencias de software, que se origina en el Instituto Tecnológico de Massachusetts [43]. Pone muy pocas restricciones en la reutilización de software, y es compatible con muchas Copyleft como GNU General Public Licence. Según GitHub y Black Duck Software, en 2015 se convirtió en la licencia más popular, por encima de variantes de la licencia GPL.

## 3.3. DISEÑO DE LA APLICACIÓN WEB

Dado que el proyecto se basa en la visualización de información, el desarrollo de la aplicación en sí misma era una parte muy importante del mismo. Para ello se siguió la

metodología formal de Diseño Guiado por Objetivos (DGO), además de otras técnicas explicadas a continuación.

### 3.3.1. BRAINSTORMING

Conocido en español como “Tormenta de ideas” se define como una herramienta de trabajo en grupo que permite generar ideas originales sobre un tema o problema concreto. Aunque hay numerosos estudios que argumentan que individualmente se generan más ideas que en grupo [44], esta técnica busca calidad en vez de cantidad, de manera que se realice un proceso de selección de las que más adapten a las necesidades del proyecto.

El principal objetivo de esta herramienta es aplazar los juicios, ya que a priori toda idea es válida y ninguna debe ser desechada, es decir, cualquier miembro del equipo puede aportar una idea relacionada con la materia en cuestión.

### 3.3.2. DISEÑO GUIADO POR OBJETIVOS

Como método formal de desarrollo de la aplicación se seguirá el Diseño Guiado por Objetivos, que todos los miembros del equipo habíamos utilizado antes, y propuesto por Alan Cooper [45]. Diseñador de software y programador, tras trabajar en Microsoft fundó su propia consultoría, donde comenzó a establecer metodologías de diseño como la que se utiliza en este proyecto, y el uso de personas como herramientas de diseño. Es una metodología que se formula la pregunta de *cómo va a interactuar el usuario con nuestro software*.

El método consta de las siguientes fases:

- 1. Investigación**

Esta fase se centra en la elaboración de un plan de investigación adecuado para el contexto del proyecto. Se realizarán una serie de entrevistas a usuarios, para identificar inquietudes y patrones, y un estudio de la competencia que nos ayude a ubicar nuestro proyecto.

- 2. Modelado**

A partir de las observaciones del apartado anterior, se definirán las personas primarias y secundarias, arquetipos de usuarios que contienen información sobre objetivos, motivaciones y comportamientos de los usuarios con el sistema.

- 3. Diseño de requisitos**

Los resultados de la etapa anterior servirán para formular los requisitos. El resultado será un conjunto de escenarios de contexto, así como una descripción de los requisitos concretos definidos en términos de contextos, acciones y objetos. Todo ello con el objetivo de facilitar el trabajo del diseño y conocer las posibles restricciones técnicas que el sistema debe incluir.



#### **4. Framework de diseño**

Representa el grueso del diseño de la aplicación. Incluye el framework de interacción, que definirá la estructura del sistema. El último paso es definir el aspecto visual de la aplicación.

Este proyecto está basado en esta metodología. Sin embargo, por cuestiones de tiempo y de lo laborioso y complejo que resultaría seguir con detalle todas las etapas y procesos, no se seguirán todas las fases de una forma estricta, sino más como una guía de desarrollo que como un método formal. Es el caso de la etapa del desarrollo del framework de diseño. El paso idílico a seguir una vez realizados los bocetos es realizar un prototipo interactivo que permita hacer evaluaciones con usuarios antes de llevar a cabo la implementación real de la aplicación. No obstante, como ya hemos mencionado anteriormente, por cuestión de tiempo fue necesario que pasáramos directamente a la implementación y lleváramos a cabo esas evaluaciones sobre la aplicación real y no sobre un prototipo.



# CAPÍTULO 4

## TECNOLOGÍAS DEL PROYECTO

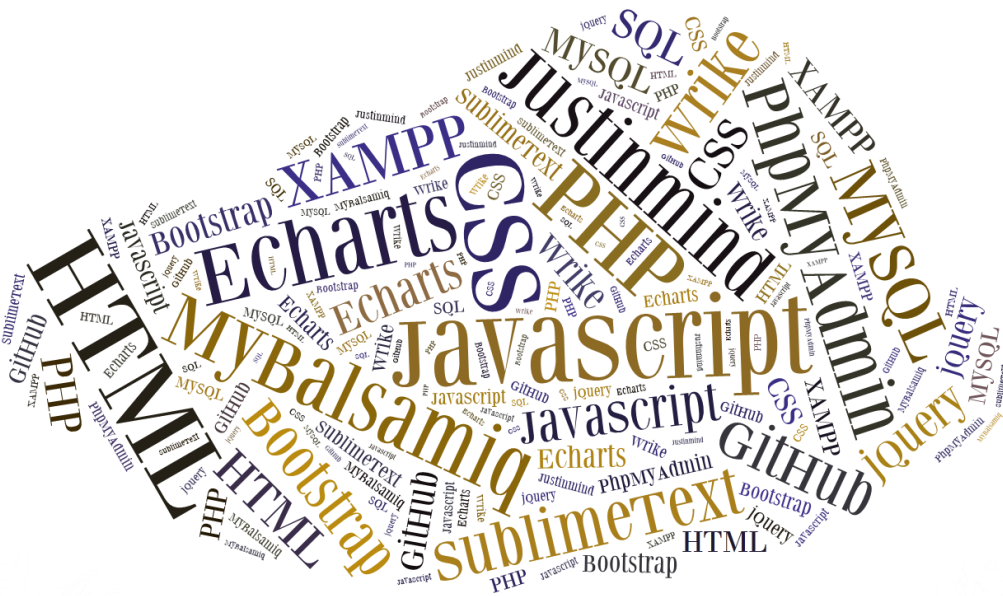


FIGURA 14. NUBE DE TECNOLOGÍAS DEL PROYECTO

En este capítulo se detallarán las tecnologías y herramientas que se han utilizado en el desarrollo de este Trabajo de Fin de Grado. Se explicará con más detalle en qué consiste la librería Echarts utilizada para trabajar con los gráficos de la aplicación, ya que es una tecnología menos popular y conocida que el resto de las que se presentan.

La siguiente tabla muestra un resumen de las principales tecnologías utilizadas durante este TFG.

Tipo	Nombre	Uso	Sección
Lenguaje	HTML	Lenguaje de marcado para la elaboración de páginas web	4.1
	CSS	Lenguaje usado para definir el estilo de un documento estructurado HTML o XML	4.2
	JavaScript	Lenguaje interpretado del lado del cliente	4.3
	PHP	Lenguaje para el desarrollo de código del lado del servidor	4.4

	SQL	Lenguaje para definir e interactuar con la Base de Datos	4.5
<b>Librería</b>	Echarts	Librería de gráficos interactivos	4.6
	jQuery	Librería JavaScript que permite la manipulación de documentos HTML	4.7
	Bootstrap	Framework basado en HTML y CSS	4.8
<b>Prototipado</b>	MyBalsamiq	Herramienta para la creación de bocetos o “mockups”	4.9
	Justinmind	Herramienta para la creación de prototipos interactivos	4.10
<b>Base de datos</b>	MySQL	Sistema gestor de la base de datos	4.11
	PhpMyAdmin	Herramienta de administración de la base de datos de la aplicación	4.12
<b>IDE</b>	XAMPP	Servidor web libre para la interpretación de páginas dinámicas	4.13
	Sublime Text	Editor de texto para la programación de la aplicación web	4.14
<b>Administración de proyectos</b>	Wrike	Plataforma web para la gestión de tareas y organización del proyecto	4.15
	GitHub	Plataforma web para el Control de versiones que aloja los desarrollos de la aplicación	4.16

TABLA 1. TECNOLOGÍAS DEL PROYECTO

## 4.1. HTML

Desarrollado en 1986 por el físico Tim Berners-Lee, HTML (*HyperText Markup Language*) es el lenguaje de marcas de texto de mayor utilización en Internet [46]. A lo largo de los años se estandariza respecto a unas normas definidas por el World Wide Web Consortium (W3C), después de que el primer estándar oficial (HTML 2.0, debido a que ni HTML ni su versión HTML+ son debidamente estandarizadas) se lleve a cabo por parte del organismo IETF (Internet Engineering Task Force). La versión definitiva del estándar actual, HTML5, se publica en octubre de 2014.

Se decidió utilizar HTML5 debido a la lista de ventajas que ofrece en el desarrollo de páginas web. Entre ellas:

- Nativo: HTML5 es nativo y, por lo tanto, es independiente de terceros y plugins.
- Canvas: Permite el uso y modificación de imágenes con mayor potencia que la etiqueta *img*, proporcionando efectos visuales.
- Proceso de Desarrollo Rápido: No hay que esperar por una compilación, actualizaciones o depuración en tiempo real.
- Simplicidad de código: permite hacer páginas ligeras que se cargarán más rápidamente, favoreciendo la usabilidad.

## 4.2. CSS

CSS (*cascading style sheets*) es un lenguaje utilizado para dar estilo a un documento estructurado escrito en HTML o XML, separando así el contenido de la presentación. El World Wide Web Consortium es, como en el caso de HTML, el encargado de formular la especificación de las hojas de estilo que servirán de estándar. La primera especificación oficial de CSS se publicó en 1995, CSS1.

CSS funciona a base de reglas, es decir, declaraciones sobre el estilo de uno o más elementos. Las hojas de estilo están compuestas por una o más de esas reglas aplicadas a un documento HTML o XML. Este sistema permite a los desarrolladores Web controlar el estilo de múltiples páginas Web al mismo tiempo, ya que cualquier cambio en el estilo marcado para un elemento en la CSS afectará a todas las páginas vinculadas a esa CSS en las que aparezca ese elemento.

## 4.3. JAVASCRIPT

Originalmente desarrollado por Brendan Eich de Netscape en 1995, JavaScript [47] es un lenguaje de programación interpretado y orientado a objetos. Se usa principalmente en su forma del lado del cliente, implementado como parte de un navegador web. Tradicionalmente se utilizaba en páginas web HTML para realizar operaciones y únicamente en el marco de la aplicación cliente, sin acceso a funciones del servidor. Sin embargo, actualmente se utiliza para enviar y recibir información del servidor junto con ayuda de otras tecnologías como AJAX. Es usada con licencia por los productos creados por Netscape Communications y entidades actuales como la Fundación Mozilla.

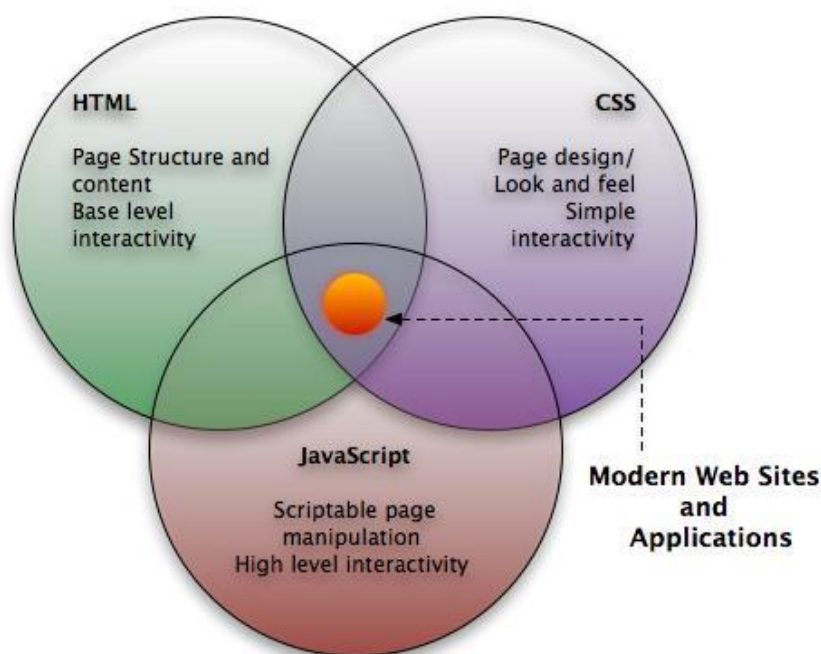


FIGURA 15. BALANCE DE TECNOLOGÍAS JAVASCRIPT, HTML Y CSS EN LA WEB

## 4.4. PHP

PHP es un lenguaje de programación [48] de código del lado del servidor, diseñado originalmente para el desarrollo web de contenido dinámico. El código es interpretado por un servidor web con un módulo de procesador de PHP que genera la página resultante. La licencia PHP es la licencia bajo la que se publica este lenguaje. De acuerdo a la Free Software Foundation es una licencia de software libre no *copyleft*. Debido a la restricción en el uso del término "PHP", no es compatible con la licencia GPL.

Todos los miembros del equipo habíamos utilizado PHP con anterioridad, y dado que está orientado al desarrollo de aplicaciones web dinámicas con acceso a información almacenada en una base de datos la decisión de utilizarlo fue unánime. Además, tiene la ventaja de poder incorporarse directamente en el documento HTML en lugar de llamar a un archivo externo que procese los datos.

## 4.5. SQL

SQL (*Structured Query Language*) es un lenguaje declarativo [49] de acceso a bases de datos cuya primera publicación hecha por ANSI y confirmada por la International Standards Organization (ISO) en 1987 está ligada al origen de las bases de datos relacionales. Sin embargo, fue Oracle quien lo introdujo como producto comercial en 1979. El álgebra y el cálculo relacional que lo caracterizan permiten efectuar consultas con el fin de recuperar información de bases de datos, así como hacer cambios en ellas.

En este proyecto se utilizará SQL para construir las consultas necesarias para acceder a la información de la base de datos y devolver al cliente de la aplicación los datos que solicite.

## 4.6. ECHARTS

Como comentamos en la sección 2.4.2.2, barajamos dos principales librerías de JavaScript con las que trabajar en nuestro proyecto. Finalmente nos decantamos por esta debido a que Chart.js solo dispone de 6 tipos de gráficos, mientras Echarts ofrece más de 11 tipos de gráficos, con 17 subtipos. Escrita en JavaScript y basada en ZRender, esta librería funciona con Chrome, Firefox, Safari, Opera e IE. Cada uno de sus gráficos está equipado con 7 componentes interactivos. Dispone de una API bien estructurada y con numerosos ejemplos que consultar.

Además, Datamatic [50], un gestor que permite la creación y visualización de gráficos basados en D3.js, diseñó una versión especial basada en Echarts con numerosos ejemplos gracias al cual pudimos hacernos una idea de qué gráficos serían más apropiados para cada uno de los análisis que pretendíamos incluir en la aplicación, sin tener que ir probando código que luego no fuera a resultarnos útil.



FIGURA 16. EJEMPLO DE GRÁFICO DE BARRAS DE ECHARTS

## 4.7. JQUERY

jQuery [51] es una librería JavaScript open-source que permite la manipulación de documentos HTML y facilita la gestión de eventos, efectos y animaciones. Es decir, hace que la programación “scripting” sea más fácil y rápida del lado del cliente.

Esta librería fue publicada por primera vez en Enero del 2006 por John Resig [52]. Es una de las librerías más empleadas dentro del mundo del diseño para la creación de proyectos digitales *responsive* o adaptativos, y además tiene algunas otras ventajas que la hacen tan popular:

- Es compatible con CSS3
- Integración con Ajax
- Es sencillo añadirla a cualquier página web. Para incluir dentro del código debemos colocar el `<script>` dentro del `<head>`
- Encadenamiento de enunciados. Permite manipular distintos elementos a la vez con una única línea de código

## 4.8. BOOTSTRAP

Bootstrap [53] es un framework creado por Twitter y basado en los últimos estándares de desarrollo web (HTML5, CSS3, JavaScript, jQuery) que agiliza la creación de la interfaz de una página web.

En este proyecto se aprovechan algunas de sus ventajas:

- Sintaxis HTML
- Diseño en malla. Permite crear una estructura de la web de manera que la distribución del contenido se vuelve flexible
- Diseño *responsive*. adapta el diseño de la página web a cualquier dispositivo

- CSS incorporado. A pesar de que se usa CSS en bastantes partes del proyecto, la inclusión de uno propio por parte de Bootstrap permite adaptar la página web a los diseños más utilizados en la web: utiliza un preprocesador llamado Less, que posee variables y funciones que aumentan la funcionalidad de CSS.
- Compatibilidad con la mayoría de navegadores

## 4.9. MYBALSAMIQ

Una de las fases más clásicas y necesarias en el desarrollo de aplicaciones es el llamado boceto o “mockup”, que básicamente se utiliza para crear una interfaz que pueda ser testeada antes de diseñar y montar. Este mockup sirve tanto para concretar detalles e ir desarrollando el diseño final, como para presentar a otros trabajadores del equipo. MyBalsamiq [54] proporciona un servicio que permite realizar esquemas de una manera sencilla. Cuando iniciamos la aplicación disponemos de una hoja en blanco a la que iremos incorporando los diferentes elementos que pasarán a formar parte de la interfaz.

Para este proyecto se utilizará MyBalsamiq como sustituto de los bocetos en papel, debido fundamentalmente a la posibilidad de compartir los esquemas. El riesgo de traspapelar folios con los bocetos, además, se eliminaba.

## 4.10. JUSTINMIND

Justinmind [55] es una de las startups españolas más sólidas [56] (afincada ahora en Silicon Valley) cuyo producto principal es una plataforma para definir y validar aplicaciones. Además da un feedback muy útil para mejorar la aplicación y seguir a la siguiente fase del desarrollo, ahorrando los costes de cambios en una versión final.

En el proyecto comenzamos a utilizar Justinmind durante la fase de diseño de la aplicación web para elaborar un prototipo interactivo, pero no pudimos más que realizar una de las funcionalidades antes de decidir abandonar el prototipado y meternos de lleno en la implementación. Aun así decidimos incluirla en esta memoria porque sí llegamos a trabajar con ella, aunque el objetivo inicial (realizar un prototipo interactivo de alta fidelidad que nos permitiera pasar a la implementación con un diseño ya afianzado) no se pudiera lograr.

## 4.11. MYSQL

MySQL [57] es un Sistema de Gestión de Bases de Datos relacional (SGBD). Inicialmente desarrollado por MySQL AB, fue comprada en 2010 por Oracle Corporation. Contrariamente a lo que ocurre con proyectos como Apache, MySQL está patrocinado por una empresa privada, que posee el copyright de la mayor parte del



código. Esto posibilita el esquema de doble licencia: GPL o versión comercial. Está desarrollado en su mayor parte en ANSI C y C++, y es una de las más populares para entornos de desarrollo web.

Utilizaremos MySQL como SGBD para almacenar la base de datos de nuestra aplicación. Este gestor aporta velocidad a la hora de realizar las operaciones, tiene un bajo coste en requisitos y el entorno es fácil de configurar e instalar.

#### 4.12. PHPMYADMIN

Otro de los motivos para elegir XAMPP es que incluye el módulo phpMyAdmin. PhpMyAdmin [58] es una herramienta escrita en PHP que permite el manejo y administración de MySQL. Su interfaz se basa íntegramente en el navegador, y actualmente puede crear y eliminar bases de datos, manipular tablas (alterar, borrar, añadir campos), ejecutar sentencias SQL, administrar privilegios de usuarios y exportar las bases de datos en varios formatos. Es un proyecto vigente desde 1998 y disponible bajo la licencia GPLv2.

En este proyecto se usará PhpMyAdmin como sistema para administrar la base de datos MySQL que contiene los datos de nuestra aplicación web.



FIGURA 17. PHPMYADMIN

#### 4.13. XAMPP

XAMPP [59] consiste fundamentalmente en el sistema de gestión de bases de datos MySQL, el servidor web Apache y los intérpretes para lenguajes PHP y Perl. XAMPP está liberado bajo la licencia GNU y actúa como un servidor web libre y capaz de interpretar páginas dinámicas. Actualmente está disponible para Microsoft Windows, GNU/Linux, Solaris y Mac OS X.

El paquete de XAMPP es de fácil instalación, y además todos los miembros del equipo habíamos trabajado con él, por lo que la elección de este entorno fue unánime.

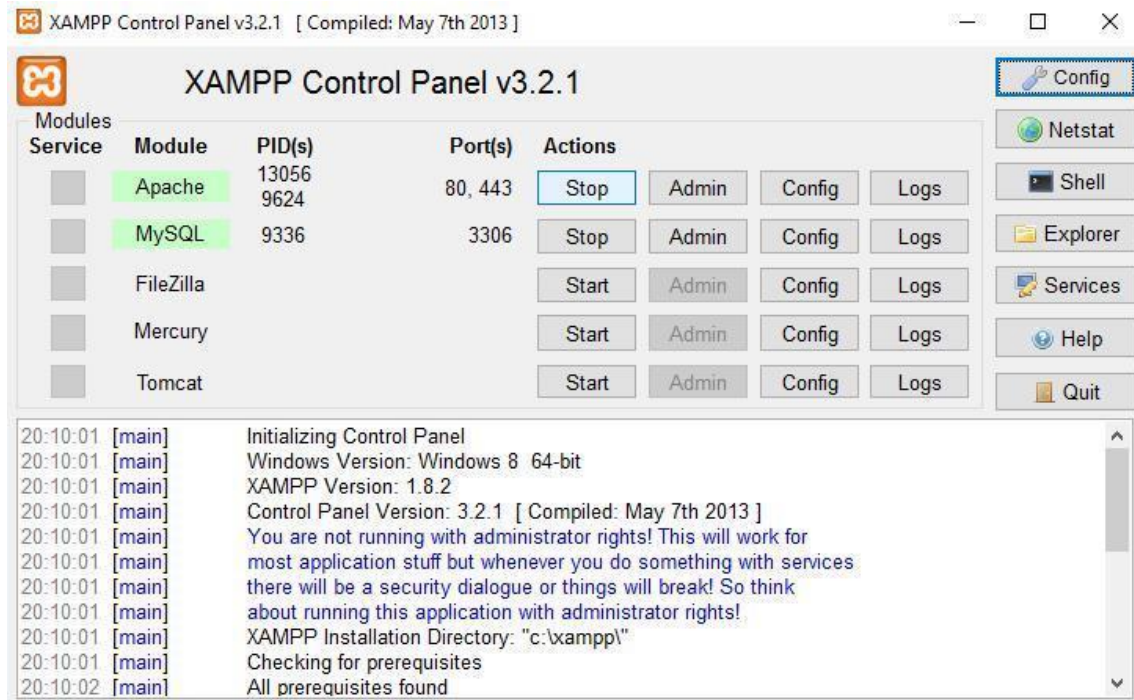


FIGURA 18. PANEL DE CONTROL DE XAMPP

#### 4.14. SUBLIME TEXT

Sublime Text [60] es un editor de código fuente escrito en C++ y Python. Se desarrolló originalmente como una extensión de Vim, la versión mejorada del editor de texto vi incluido en todos los sistemas UNIX, pero con el tiempo adquirió una identidad propia. Soporta un gran número de lenguajes, su interfaz es limpia e intuitiva y soporta el uso de snippets, plugins y sistemas de construcción de código (*Build Systems*).

Sublime Text puede ser descargado y evaluado de forma gratuita. Sin embargo, no es software de código abierto, ya que se debe obtener una licencia si se desea hacer un uso continuado del mismo, aunque la versión de prueba no tiene fecha de caducidad y es totalmente funcional. En este proyecto se utilizará Sublime Text para llevar a cabo la edición y programación en PHP de la aplicación web, ya que no se requiere un entorno de desarrollo específico para ello.

#### 4.15. WRIKE

Wrike [61] es un software para la colaboración online. Permite que sus usuarios ajusten sus planes de proyectos, prioricen tareas, estén al tanto de la planificación y colaboren online con sus compañeros de equipo.

Este software está disponible en tres versiones, y todas incorporan un flujo que informa a los usuarios sobre las acciones llevadas a cabo por otros usuarios pertenecientes al grupo de trabajo. La pantalla principal incluye una barra lateral donde podemos ver las distintas categorías de gestión y el acceso a los Proyectos, además de un acceso a los

Filtros Rápidos. En la sección de Mi Trabajo se muestra el listado con las tareas asignadas (pendientes, completadas) organizadas en función del día actual, la semana actual, la próxima semana o más adelante. Después tenemos el Panel de Información o dashboard, con las tareas asignadas y las tareas en espera, entre otras. El Flujo muestra las últimas actualizaciones relativas a cada una de las tareas activas, con un resumen de los comentarios que cada uno de los miembros del equipo ha podido hacer.

En resumen, es una herramienta intuitiva y que permite hacer un seguimiento completo de la evolución del trabajo dentro de cada proyecto de manera sencilla y práctica.

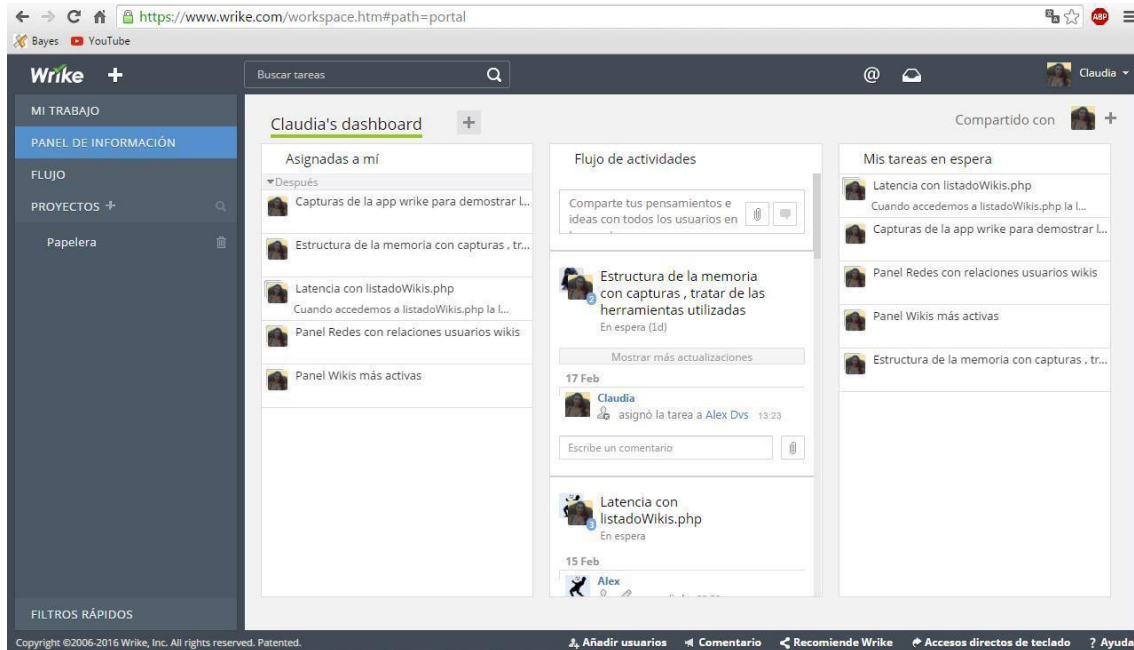


FIGURA 19. PANEL DE GESTIÓN DE PROYECTOS DE WRIKE

## 4.16. GITHUB

Aunque hasta casi principios de Mayo cada uno de los miembros del equipo trabajamos en local, este sistema no resulta práctico ni adecuado cuando el desarrollo alcanza un cierto volumen, por lo que, además de seguir gestionando las tareas mediante Wrike, decidimos abrir un proyecto en Github.

Github [62] es un servicio de alojamiento de repositorios de software con sistema de control de versiones Git. Además, ofrece varias herramientas útiles para el trabajo en equipo y colaboración, como una wiki para el mantenimiento de las distintas versiones, un sistema de seguimiento de problemas, una herramienta que permite la revisión de código donde se pueden incluir anotaciones, así como un visor de ramas para comparar los progresos realizados.



# CAPÍTULO 5

## DISEÑO DE LA APLICACIÓN

En este capítulo se explicará en profundidad el proceso de desarrollo del diseño de la aplicación. Cabe mencionar que este desarrollo se llevó a cabo en paralelo a la obtención de datos de Wikia, como se explicará en el capítulo 7, y que, a medida que recogimos más información, las funcionalidades de la misma aumentaron ligeramente. Es por esto que la aplicación final alberga más elementos y funcionalidades que los prototipos que se detallan al final de este capítulo.

### 5.1. INTRODUCCIÓN

El primer contacto con la aplicación requería una parte de investigación sobre el funcionamiento de Wikia, dado que habíamos decidido centrarnos en esas comunidades. Por tanto, era necesario saber exactamente qué información concreta podíamos obtener de esta comunidad, dado que dependiendo de las características de dicha información el desarrollo de la aplicación tomaría un rumbo u otro (ver sección 2.2.1).

Tras varias sesiones de brainstorming entre los miembros del equipo se llegó a posibles implementaciones de la aplicación que resultaran útiles para mostrar el tipo de análisis que teníamos en mente.

### 5.2. DISEÑO GUIADO POR OBJETIVOS

La primera etapa de esta sección consiste en desarrollar el Diseño Guiado por Objetivos, iterando las 4 etapas de este proceso como se definió en la sección 3.3.2, que detallamos a continuación.

#### 5.2.1. INVESTIGACIÓN

En esta sección se detallará el proceso de investigación previo realizado para concretar la funcionalidad de la aplicación, así como su diseño. En él vamos a recabar información en base a entrevistas con usuarios y al estudio de mercado cubierto en el Capítulo 2.

Para las entrevistas escogimos a un alumno de la facultad y dos personas ajenas a la misma, de distintas edades y trabajos, pero con interés en plataformas colaborativas. Los resultados de estas entrevistas serán expuestos con más detalle a partir de la sección 5.3. De todos los puntos que hablamos con ellos sacamos en claro lo siguiente:

- En primer lugar, a todos los candidatos les parece interesante y útil una aplicación en la que se pueda consultar información sobre proyectos colaborativos, y están de acuerdo en que ellos harían uso de la aplicación.
- En segundo lugar, ninguno de los usuarios conoce ni ha oído hablar de una herramienta similar a la que les presentamos. Uno de los entrevistados reconocía que él utilizaba por motivos de trabajo herramientas para mostrar información tipo dashboards pero en ningún caso relacionadas con comunidades colaborativas.
- Por último, aunque dos de los tres usuarios estaban familiarizados con Wikia, todos coincidían en que si se lograba incluir otras comunidades como StackOverflow o Wikipedia, los posibles usos de la herramienta crecían exponencialmente.

El estudio de la competencia se centra en el análisis de cómo los usuarios utilizan y perciben un producto que ya existe, y ayuda a aprender cómo el sistema proporciona soporte en la realización de una determinada función y a conocer cuáles son los principales problemas de usabilidad que presentan, con el fin de así poder evitarlos. La información que se pretende extraer de este tipo de estudio es fundamentalmente:

- Qué características de un producto resultan útiles a los usuarios
- Qué problemas encuentran los usuarios al utilizar dicho producto

Además, otra fuente de información que hemos considerado importante de cara al diseño es el conjunto de revisiones de esos productos hechos por terceras partes (blogs, artículos, etc). Por ello lo hemos tenido en cuenta a la hora de analizar algunas de las herramientas de visualización descritas anteriormente (ver sección 2.4).

#### Recomendaciones de acción:

Tras el análisis de estas herramientas se puede extraer información muy provechosa sobre las fortalezas, debilidades y oportunidades. Está claro que el aspecto visual a la hora de proporcionar análisis de una cierta información (del tipo que sea) es muy importante para que la herramienta resulte útil a los usuarios. La claridad de los gráficos que se incorporen es fundamental para que la información no resulte confusa o engañosa, es decir, hay que **elegir muy bien el tipo de gráfico a utilizar dependiendo de lo que se quiera mostrar**. Parece también claro que una interfaz simple y clara

facilita mucho el uso de una aplicación, evitando distracciones que le quiten “protagonismo” al objetivo final, la comprensión de la información.

## 5.2.2. MODELADO

A partir de las conclusiones obtenidas en la etapa de investigación, se definirán las personas primarias y secundarias. La *persona* es una descripción detallada del individuo imaginario que representa a un conjunto de posibles usuarios a los que va destinada la aplicación. Las *personas* son una herramienta más en el diseño de la aplicación, y su objetivo es centrar ese diseño en ese tipo de persona para lograr los objetivos propuestos a la hora de utilizar el sistema. La estrategia de procesamiento seguida se basa en el modelo *top-down*. Este modelo con frecuencia se diseña con ayuda de “cajas negras” que hacen más sencillo cumplir los requisitos, aunque no expliquen en detalle los componentes individuales. En él se formula un resumen del sistema, y cada parte de ese sistema se refina a su vez con mayor detalle, hasta que la especificación completa es lo suficientemente detallada para validar el modelo.

### 5.2.2.1. Identificación de categorías de usuarios

Utilizamos la estrategia de división por roles de usuario. Inicialmente encontramos tres candidatos a formar parte de un rol real, siendo los dos primeros las personas primarias:

- Usuarios de **Wikia**
- Usuarios interesados en **producciones colaborativas**
- Usuarios de internet, que no participen necesariamente de una forma activa en comunidades colaborativas pero que sí sean consumidores de toda la información ofrecida

### 5.2.2.2. Procesamiento de los datos y esqueletos

En esta fase definiremos el tipo de persona que interactuará con nuestra aplicación.

- **Usuarios de Wikia** (15 años en adelante). Colaboradores activos y consumidores habituales de Wikia, como administradores que quieren conocer información sobre otras wikis (visitas, logros habilitados, número de ediciones) con el objetivo de mejorar su propia página, pudiendo comparar esa información con la de su propia comunidad. Utilizan frecuentemente la aplicación
- **Usuarios interesados en producciones colaborativas** (25 años en adelante). Principalmente gente interesada en el mundo de las comunidades colaborativas



(no necesariamente Wikia) y con interés estadístico y analítico (perfiles de usuarios, idiomas en los que más se participa, evolución en el tiempo de las colaboraciones) Utilizan frecuentemente la aplicación

- **Usuarios de internet** (15 años en adelante). Usuarios que, puntualmente, puedan necesitar acceder a este tipo de información por motivos, o bien de mera curiosidad, o bien alguna investigación concreta. Utilizan de forma esporádica la aplicación

En las siguientes figuras se muestran una serie de personas ficticias que podrían representar el tipo de usuario al que va dirigida la aplicación.

**Nombre:** Felipe

**Edad:** 19 años

**Ocupación e intereses:** estudiante de informática en la Universidad Complutense de Madrid. Apasionado de la Playstation y miembro activo de Wikia desde hace 2 años. Colabora en varias comunidades sobre algunos de sus videojuegos favoritos. Le gustaría poder llevar un control sobre cómo afecta el contenido de esas páginas al número de colaboraciones y las preferencias que tienen los usuarios.

**Cita:** *"Me gustaría saber qué repercusión tienen los cambios que introduzco en mi página"*



**Nombre:** Sara

**Edad:** 32 años

**Ocupación e intereses:** licenciada en Ingeniería Informática por la Universidad Autónoma de Madrid. Comprometida con proyectos Open Data, y con experiencia en el desarrollo de herramientas web para la producción *peer-to-peer*. Sus intereses incluyen los movimientos sociales y el desarrollo de herramientas de visualización para el análisis de comportamiento de redes sociales.

**Cita:** *"Me gustaría utilizar herramientas que ofrezcan una visualización interactiva sobre datos de plataformas online"*





<p><b>Nombre:</b> Juan Carlos</p> <p><b>Edad:</b> 36 años</p> <p><b>Ocupación e intereses:</b> licenciado en matemáticas e investigador en la Universidad de Málaga. Desarrolla herramientas para el análisis de proyectos de software libre. Ofrece con frecuencia seminarios y charlas sobre métodos útiles en relación al análisis estadístico. Está interesado en el estudio sobre cómo se mantienen y gestionan comunidades online cuyo propósito es la generación de conocimiento compartido.</p> <p><b>Cita:</b> “Me gustaría disponer de más herramientas que faciliten el análisis de algunos proyectos digitales”</p>	
---	--

FIGURA 20. FICHAS FICTICIAS DE PERFILES DE USUARIOS DE CHARTSUP

### 5.2.2.3. Priorización de esqueletos

La conclusión que se puede obtener es que los usuarios de internet que sólo necesiten consultar información muy de vez en cuando utilizando nuestra aplicación no la encontrarán realmente útil, ya que su uso se verá limitado a unas pocas interacciones. Por lo tanto, no se han incluido como usuarios finales de la aplicación.

## 5.2.3. DEFINICIÓN DE REQUISITOS

A partir de los resultados obtenidos en las secciones anteriores, formulamos los requisitos de nuestra aplicación. El objetivo es situarnos en los escenarios reales que pudieran surgir. En cada uno de ellos se especificarán los requisitos necesarios para conseguir cada uno de los objetivos o solventar los problemas. En este apartado conviene diferenciar los términos de acción, objeto y contexto con los que vamos a trabajar. La acción se define como la actividad que requiere una solución. El objeto es el sujeto principal del escenario que está siendo tratado. Y por último, el contexto especifica el objetivo final que persigue el requisito.

### Escenario I

Oscar lleva la administración de una página wiki sobre una serie de suspense desde hace 2 meses, pero no está seguro de si lo está gestionando correctamente. Por ello decide comparar su página con otra de un perfil similar. Cuando compara las dos páginas se da cuenta que los usuarios son algo más activos en su comunidad, pero el número de ediciones es mayor en la otra debido a que en esa los usuarios burócratas participan más.

Requisitos:

Comparar (**acción**) los aspectos más importantes de dos comunidades, tales como la actividad de sus usuarios, la media de ediciones de cada wiki o las páginas por wiki (**objeto**) con el fin de observar diferencias y tratar de mejorar alguna de ellas (**contexto**).

**Escenario II**

Alberto es investigador del Departamento de Estadística de la Universidad de la Rioja y tiene interés en analizar qué ritmo de publicaciones llevan los usuarios más activos de Wiki, y si además esos usuarios *core* participan en una o varias comunidades. Para ello, a través de la sección Global User Wiki Statics de la aplicación accede al análisis del perfil de uno de ellos.

Requisitos:

Visualizar (**acción**) los distintos perfiles de usuarios más importantes (**objeto**) para poder así analizar su comportamiento en el conjunto de varias wikis (**contexto**).

**Escenario III**

Miguel tiene que presentar un trabajo en el instituto sobre el fenómeno fan y popularidad de un tema que le interese. Como es un gran fan de Star Wars, y aprovechando la última entrega de la saga, decide hacer un pequeño estudio sobre cómo de popular es en Wikia dicha comunidad. Para ello, consulta en la aplicación toda la información relativa a esa wiki. Encuentra los usuarios más activos, la media de ediciones, cuáles son los usuarios que más aportan a la comunidad y cómo se relacionan con otras páginas de ciencia ficción.

Requisitos:

Visualizar (**acción**) los distintos gráficos y tablas disponibles de una wiki (**objeto**) con el objetivo de obtener información de un determinado tema (**contexto**).

## 5.2.4. FRAMEWORK DE DISEÑO

El framework de diseño define la estructura general de la interacción con el usuario. En él establecemos la organización de los elementos funcionales de la interfaz y su comportamiento. Es decir, vamos a estudiar el framework de interacción.

En esta fase vamos a especificar tanto la estructura a bajo nivel y la organización de las pantallas, como el flujo y comportamiento de nuestra aplicación. Para ello, definimos un proceso de varias etapas, partiendo primero de la descripción de los escenarios, para luego revisar el agrupamiento de las funcionalidades y realizar los bocetos que facilitarán la detección de posibles fallos o mejoras en el sistema.

#### 5.2.4.1. Definir el factor de forma, la postura y los métodos de entrada

El factor de la forma tiene en cuenta el **contexto en que se va a presentar la información disponible en la aplicación**. Así, por ejemplo, uno de los elementos principales a considerar a la hora de diseñar ChartsUp es que se trata de una aplicación web que se va a utilizar en interiores, ejecutándose en cualquier navegador web moderno, y cuyos componentes (gráficos, paneles, pestañas o textos) deben ser claros y que no resulten ajenos para el usuario.

En cuanto a la postura, se trata de una aplicación web que proporciona mucha información organizada en distintas pestañas y, en general, aunque la consulta sea breve, requiere algo de tiempo comprender las gráficas que se muestran. Es decir, el usuario utilizará el sistema durante periodos largos de tiempo, cuando quiera comparar varias wikis o consultar datos sobre una de ellas. Las opciones que permiten modificar el aspecto visual de las gráficas son iconos intuitivos y semejantes entre sí, y que proporcionan una interfaz clara. La interfaz es limpia y sin distracciones que alejen al usuario del principal objetivo de la aplicación, que es poder visualizar información útil en un entorno atractivo e intuitivo.

Los métodos de entrada elegidos, relacionados con el factor de forma, la postura y las aptitudes de las personas primarias son la pantalla del ordenador, por tratarse de una aplicación web, y ratón que permita navegar y desplazarse con facilidad por las distintas páginas.

#### 5.2.4.2. Definir los elementos y datos funcionales

En esta sección definiremos los elementos de la interfaz que representarán a los requisitos de datos y funciones que hemos obtenido en la definición previa de los escenarios de contexto. Para ello, analizaremos de manera individual cada uno de los esos escenarios en busca de elementos de datos, atributos y relaciones que se establezcan entre ellos.

Escenario I: Como elemento de datos principal tenemos la comparación de dos o más comunidades de Wikia, que se relaciona con las funcionalidades de actividad de usuarios y estadísticas generales.

Escenario II: En este caso, el elemento de datos definido es obtener información sobre uno de los usuarios de una wiki, cuyo atributo es el usuario y que se relaciona con otras comunidades en las que realiza aportes, sus últimos logros y las ediciones totales.

Escenario III: Aquí el elemento de dato, consultar una de las wikis para obtener su información, está relacionado con el tipo de usuario que participa en ella y las estadísticas generales de ediciones, páginas y ficheros.

Una vez definido el catálogo, la traducción de los requisitos anteriores a elementos funcionales nos ha proporcionado los siguientes componentes de la interfaz:

- La opción de búsqueda de wikis es una de las más importantes de la aplicación. Sin embargo, debido al gran volumen de comunidades disponibles, y teniendo en cuenta los principios de Nielsen (facilidad de aprendizaje, *cómo de fácil es para un usuario novato realizar una tarea en el sistema*; y eficiencia, *cómo de rápido puede realizar una tarea un usuario experto*) se decidió englobar esa búsqueda en dos categorías distintas: agrupadas por letra, y en una tabla de filtrado que muestra todos los resultados que coincidan con la palabra incluida.
- El comparador es otra parte fundamental y de gran potencial dentro del sistema. Así, la consulta de la información para dos o más wikis dispone de un apartado bien diferenciado dentro del panel de navegación, que da acceso al formulario. La interfaz es obvia y fácil de recordar en caso de haber pasado tiempo sin usarse.
- La información relativa a los usuarios será otro componente independiente, donde se muestren los más activos junto con acceso a sus perfiles, aplicando el principio de usabilidad **menos es más**: las interfaces sencillas suelen resultar más usables, y tratar de incluir todas las funcionalidades posibles en una única interfaz hace que el sistema sea complejo.

#### 5.2.4.3. Determinar los grupos funcionales

De forma más jerárquica, los elementos funcionales definidos en el apartado anterior quedarán agrupados de la forma que se explica a continuación:

1. Los elementos que requieren más espacio visual dentro de nuestra interfaz son búsqueda (*Search*), comparador (*Compare*) y estadísticas de usuarios (*User Statics*), ya que son las acciones más frecuentes. Por ello, se mantendrá una sección lateral en la aplicación en la que aparezcan enlaces a los elementos que corresponden a las funcionalidades anteriores. Estos tres elementos estarán fijos, de forma que resulte sencillo navegar por ellos y cambiar de uno a otro de una forma rápida y sin tener que estar volviendo continuamente a una página inicial.
2. Los enlaces de *Compare* y *User Statics* llevan directamente a las páginas de sus respectivas secciones. Pero el enlace de Búsqueda, sin embargo, estará a su vez dividido en función de las distintas categorías mencionadas anteriormente. Para facilitar la búsqueda se proporcionan dos opciones:

- Búsqueda alfabética (*Alphabetic Wiki Search*), que accederá a una pestaña en la que se agruparán por letra todas las comunidades disponibles, y
  - Búsqueda por nombre mediante una tabla de filtrado (*Filter Wiki Table Search*), que mostrará una tabla en otra pestaña en la que aparecerán todas las wikis por orden alfabético, y que incluirá una barra de búsqueda que agilice el proceso.
3. La organización de las pestañas y contenedores está pensada para que la interacción con el sistema sea fluida, de tal forma que intentamos minimizar iconos para no llenar la interfaz de botones ni textos que den sensación de sobrecarga de información. De esta forma, cada enlace descrito en el punto anterior marca el acceso a un contenedor independiente.
4. La interfaz no requiere de un orden concreto para usar la aplicación. Cada una de las funcionalidades detalladas tiene un flujo distinto, dependiendo exclusivamente de la tarea que el usuario quiera realizar.

#### 5.2.4.4. Boceto del framework de interacción

Con la información recogida hasta el momento, pasamos al desarrollo del primer boceto de la aplicación. Como ya mencionamos en la sección 4.9, la herramienta utilizada para el desarrollo de mockups es MyBalsamiq. No utilizamos en ningún momento bocetos en papel porque desde el principio tuvimos claro a grandes rasgos qué aspecto queríamos que tuviera la aplicación.

#### Escenarios Key Path y validación

Los escenarios *key path* son los escenarios que muestran cómo interactúa una persona con la interfaz diseñada en el framework de interacción. Es decir, son la continuación natural de los escenarios de contexto descritos anteriormente, pero esta vez mostrando qué haría una persona con los elementos de datos de la interfaz.

En este apartado se pondrán a prueba dichos escenarios, y utilizaremos algunas preguntas que podrían resultar lógicas para definir la respuesta de la aplicación y así tratar de mejorar su diseño.

**Escenario I:** Comparar (acción) los aspectos más importantes de dos comunidades (objeto) con el fin de observar diferencias y tratar de mejorar alguna de ellas (contexto).  
**¿Qué ocurre si una persona accede por primera vez a la aplicación y quiere comparar dos wikis?**

Cuando entramos en nuestra aplicación apareceremos en la pantalla de bienvenida. En el menú lateral estarían las opciones principales: *Buscador*, *Comparador*, *Usuarios* y *Sobre nosotros*.

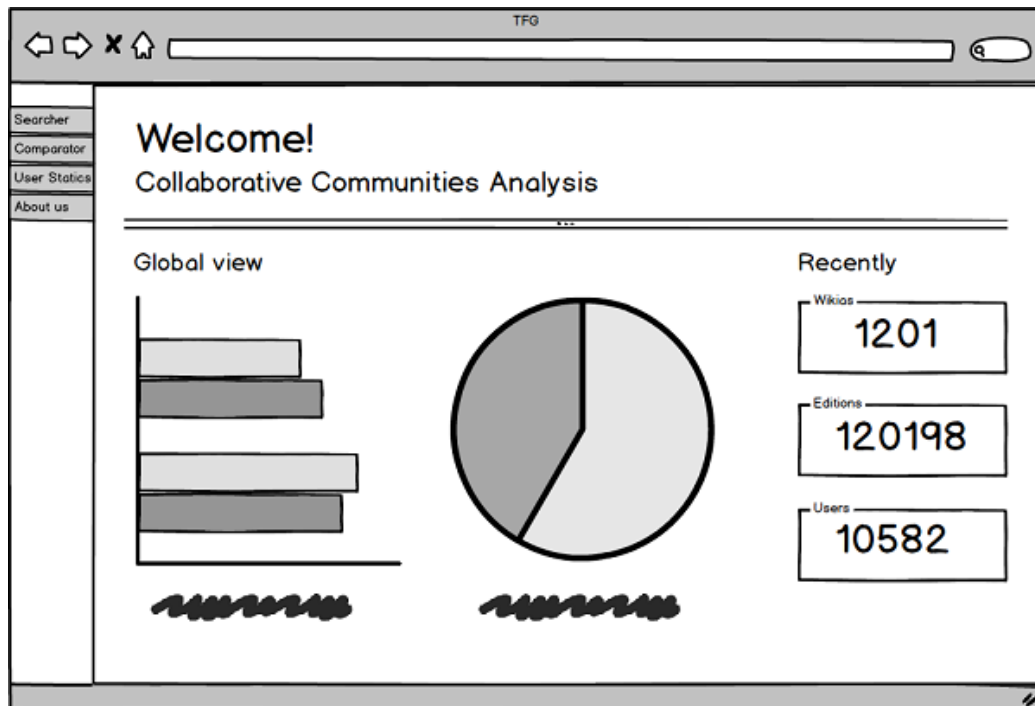


FIGURA 21. BOCETO DE LA PÁGINA DE BIENVENIDA DE LA APLICACIÓN

Nada más clicar el enlace del comparador accederíamos a la pestaña donde se situarían los cuatro desplegados en los que podremos buscar el nombre de las comunidades que queramos comparar. Después de elegirlos, el botón *compare* redirigiría a otra página con el mismo formato en la que se muestra toda la información relativa a esas dos wikis.

The wireframe shows a web browser window with a title bar containing navigation icons and a search bar. The page has a sidebar with links: Searcher, Comparator, User Statics, and About us. The main content area is titled 'Wikia comparator' and includes a message: 'Please select at least two diferent wikias'. There are four search input fields labeled 'Option 1', 'Option 2', 'Option 3', and 'Option 4'. Each field has a 'Search' button. Below the search fields is a large 'COMPARE' button. The page is styled with a clean, modern look using a light gray color scheme.

FIGURA 22. BOCETO DEL ACCESO AL COMPARADOR



FIGURA 23. BOCETO DE LA PÁGINA DEL COMPARADOR

Navegaríamos por las distintas pestañas, cada una para un tipo de información, y en cualquier momento podríamos pulsar el botón atrás del navegador y reaparecer en el formulario para comparar otras.

#### Hallazgos y recomendaciones:

Dentro del comparador, los despleables no son la mejor opción de búsqueda, dado el elevado número de comunidades disponibles. El usuario tendría que recorrer con un *scroll* hasta encontrar su wiki. Si hay un error al clicar sobre una que no nos interesa (la de encima de la que buscamos, por ejemplo) tiene que repetir el proceso. Por ello decidimos que la mejor opción era añadir a este botón una sencilla barra de búsqueda que permita encontrar la página con mucha más rapidez.

¿Qué ocurre si no seleccionamos ninguna e intentamos comparar? Si el usuario por accidente no seleccionara ninguna wiki se quedaría esperando a que la aplicación reaccionara. Se nos ocurrieron dos soluciones alternativas: o bien deshabilitar el botón de Compare, o bien incluir en los dos despleables primeros un asterisco indicando que esos campos son obligatorios.

Otra posible mejora estaba relacionada con la pantalla de bienvenida. Una vez accedemos a cualquiera de los enlaces, no hay ninguna manera de volver a acceder a la visión global del inicio, sólo a través del botón atrás del navegador. Para facilitar la navegación decidimos incluir en el menú lateral un enlace *First steps*, que redirija a la pestaña de bienvenida.

**Escenario II:** Visualizar (acción) los distintos perfiles de usuarios más importantes (objeto) para poder así analizar su comportamiento en el conjunto de varias comunidades Wikia (contexto).

**¿Cómo se consulta la información de los usuarios registrados?**

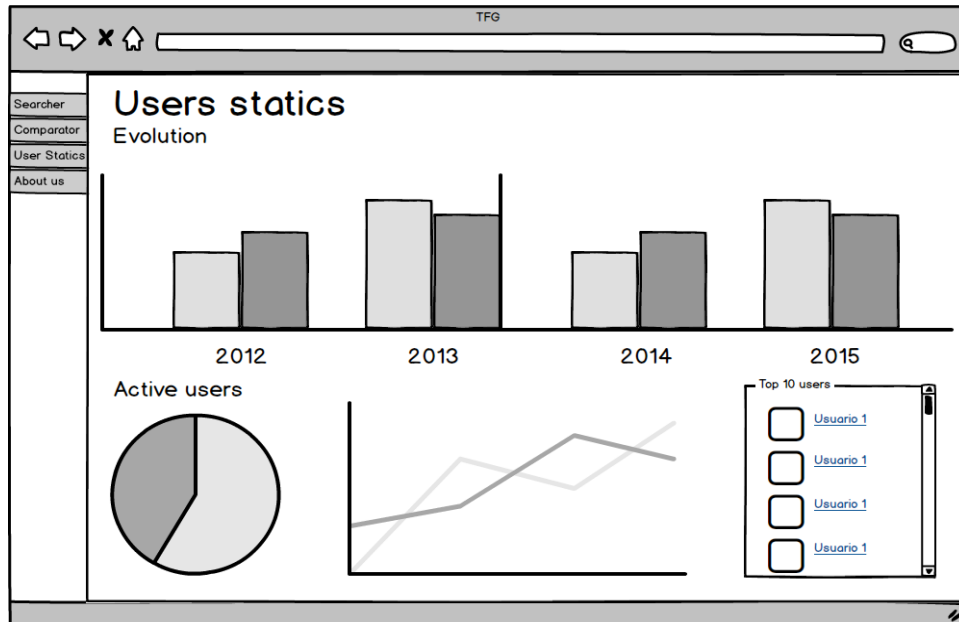


FIGURA 24. BOCETO DE LA PÁGINA DE GLOBAL USERS

De nuevo desde el menú lateral encontraríamos la opción para acceder a la pestaña de Estadísticas de usuario. A través de ella navegaríamos por las distintas gráficas y tablas de información. Además, si nos interesa el perfil de un usuario en concreto se puede buscar en una de las tablas y acceder a su perfil.

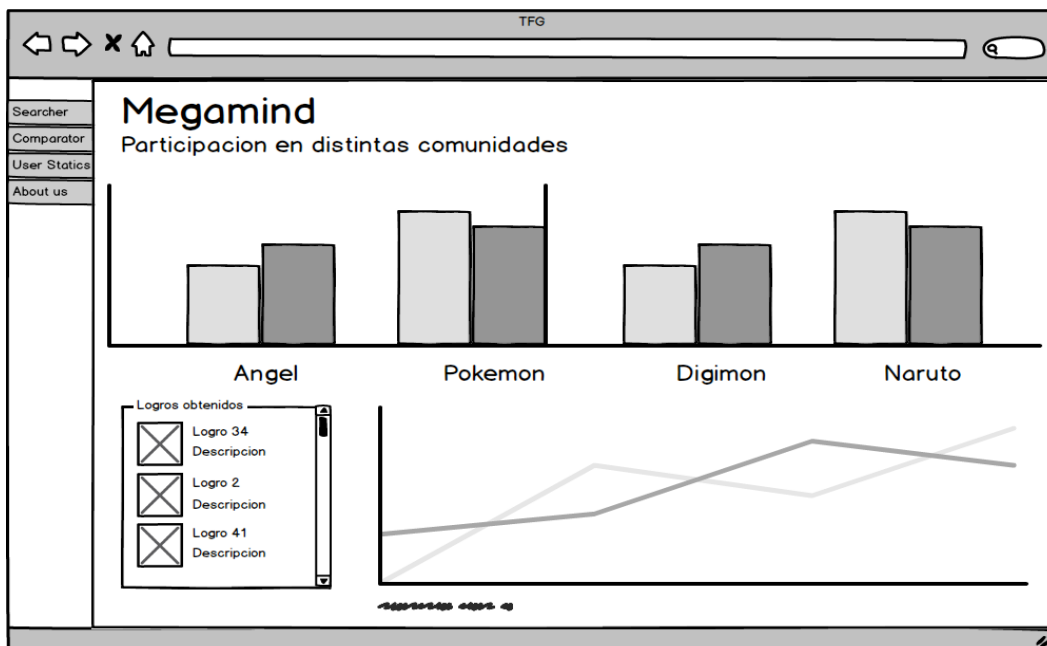


FIGURA 25. BOCETO DEL PERFIL DE UN USUARIO DE WIKIA



### Hallazgos y recomendaciones:

Tal y como planteamos inicialmente en el boceto esta sección se centraba más en los usuarios importantes, pero ¿qué pasa si se quiere consultar el perfil de un usuario que, por ejemplo, no realiza más que unos pocos aportes? ¿Cómo se accede a la información de ese usuario? Esta acción no se contemplaba, no existía ningún elemento que permitiera realizar una consulta fuera del “*top* usuarios”. Por este motivo agregamos una tabla, similar a la de búsqueda de wikis, en la que aparecen todos los usuarios, con opción de buscar por alias en la parte superior.

**Escenario III:** Visualizar (acción) los distintos gráficos y tablas disponibles de una wikia (objeto) con el objetivo de obtener información de un determinado tema (contexto).

### **¿Cómo puede acceder un usuario a la información de una wiki en particular?**

Desde la pantalla de inicio accederíamos a través del menú lateral al buscador. En ese momento aparece una ventana en la que se elige el tipo de búsqueda que queremos realizar.

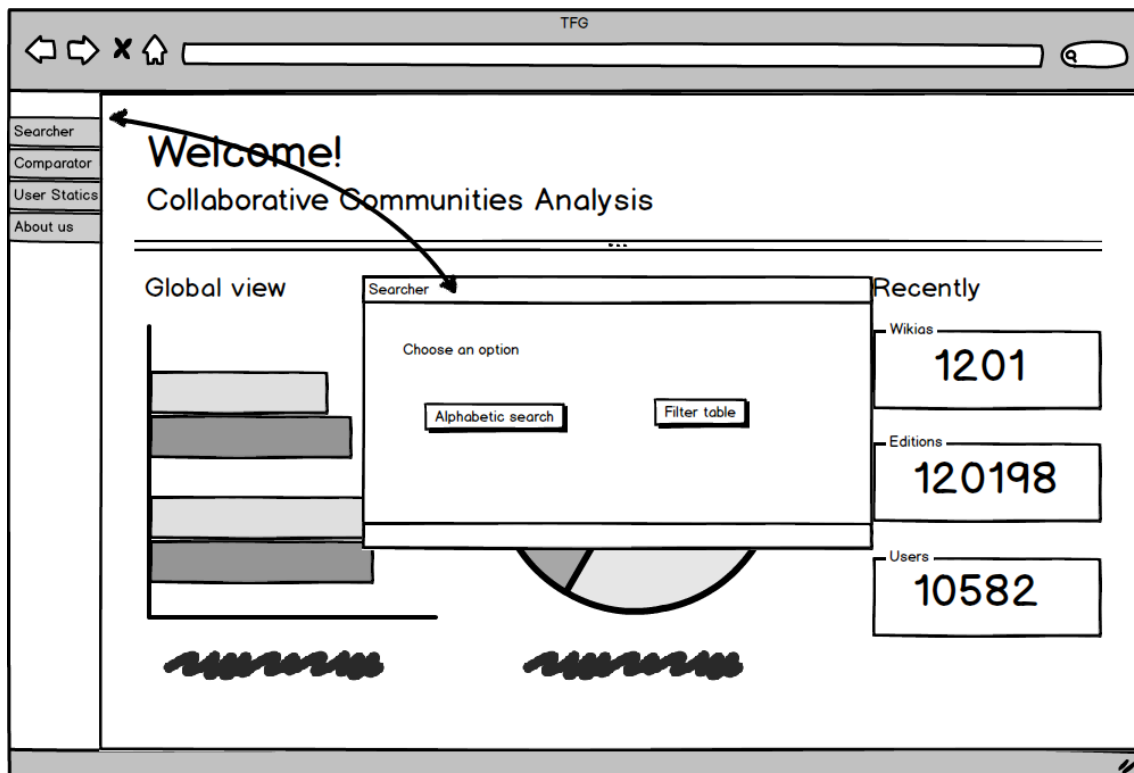
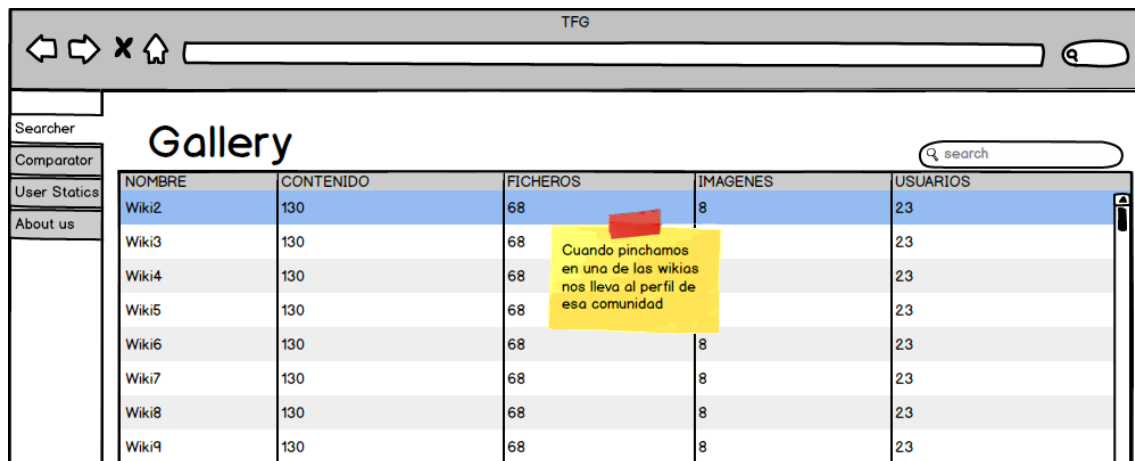


FIGURA 26. BOCETO DEL ACCESO AL BUSCADOR DE WIKIS

Si sabemos qué wiki queremos buscar, pinchamos en el botón de Filter table y la aplicación nos llevaría a la tabla donde aparecen por orden todas las comunidades registradas hasta el momento. En la barra de búsqueda superior escribiríamos el nombre de la wiki, de manera que la encontraríamos rápidamente. Luego simplemente accederíamos a través del nombre a todas sus estadísticas e información.



TFG

Searcher  
Comparator  
User Statics  
About us

## Gallery

search

NOMBRE	CONTENIDO	FICHeros	IMAGENES	USUARIOS
Wiki2	130	68	8	23
Wiki3	130	68		23
Wiki4	130	68		23
Wiki5	130	68		23
Wiki6	130	68	8	23
Wiki7	130	68	8	23
Wiki8	130	68	8	23
Wiki9	130	68	8	23

Cuando pinchamos en una de las wikias nos lleva al perfil de esa comunidad

FIGURA 27. BOCETO DE LA TABLA DE BÚSQUEDA DE WIKIS

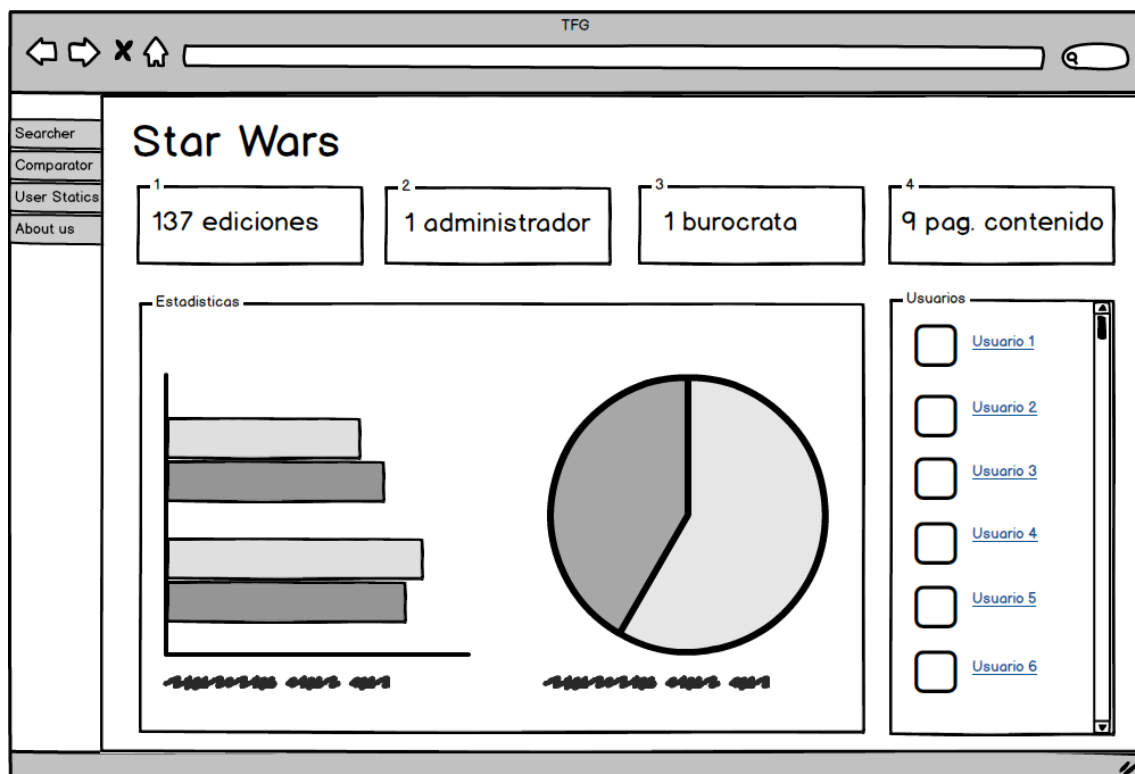


FIGURA 28. BOCETO DEL PERFIL DE UNA WIKI

### Hallazgos y recomendaciones:

No parecía muy lógico que fuera necesaria una ventana emergente para elegir el tipo de búsqueda. Es lento y más costoso, por lo que la opción del pop-up se descartó

rápidamente. En su lugar, incluimos las dos opciones de búsqueda dentro del propio menú lateral, en forma de desplegable.

Llegados a este punto, el paso lógico siguiente era construir un prototipo interactivo de alta fidelidad que permitiera hacer evaluaciones para ahorrar tanto tiempo como esfuerzo en la implementación, ya que requiere mucho más trabajo realizar modificaciones sobre una aplicación ya desarrollada que sobre un prototipo. Sin embargo, como ya explicamos en la planificación (ver sección 3.3.2), tuvimos que saltarnos este paso (sólo realizamos a alto nivel una de las funcionalidades) y comenzar directamente a desarrollar, sin haber tenido en cuenta más opiniones y mejoras que las que planteamos nosotros mismos al examinar los bocetos.

## Principios de diseño

A continuación detallaremos los principios de diseño [63] en los que hemos basado este desarrollo:

**Principio de consistencia interna:** La consistencia trata de garantizar que aquellos controles que tienen una funcionalidad parecida deben así mismo tener una apariencia similar, con el fin de que el usuario se adapte más fácilmente al entorno de la interfaz y vea las potencialidades de sus elementos.

En nuestro caso, el tipo de letra y disposición de los paneles es similar entre unas páginas y otras. Además, la gama de colores utilizada es la misma para todos los gráficos (salvo que se necesitara resaltar alguna categoría en particular).

**Principio de Gestión del estado visible:** Para que el usuario pueda mantenerse en sintonía con el sistema hay que permitirle establecer flujos mentales mientras interactúa con la aplicación. El usuario de la aplicación tiene que saber dónde se encuentra en cada momento, y tiene que poder volver a donde desee sin que resulte complejo. En nuestra herramienta se consigue manteniendo todos los enlaces principales fijos en un panel lateral, además del botón de atrás incluido en los navegadores, que permite volver al paso anterior de una manera intuitiva.

**Principio de Visibilidad y feedback:** Este principio tiene que ver con cómo la interfaz utiliza diferentes mecanismos para transmitir su estado actual y las posibles acciones en un determinado momento.

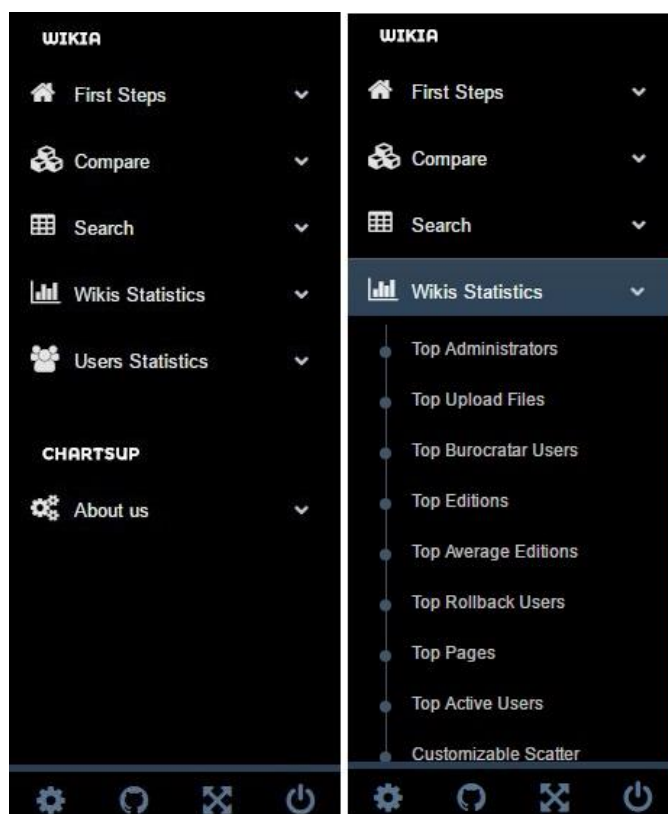


FIGURA 29. ENLACES OCULTOS/VISIBLES DEL MENÚ PRINCIPAL

1. **Visibilidad progresiva y jerárquica:** Cuando un sistema ofrece múltiples funcionalidades, organizar la interfaz es una tarea compleja ya que, en vez de ofrecer todas las funcionalidades disponibles a la vez, el sistema debe presentar únicamente las que son relevantes en un determinado momento. Es decir, tratar de mostrar todas las potencialidades de un golpe suele provocar una interfaz saturada y poco amistosa. Por este motivo es conveniente establecer jerarquías visuales. Nuestra aplicación desplegará las opciones de cada uno de los enlaces principales únicamente si hemos pinchado previamente en uno de ellos. No tiene sentido mostrar la opción de realizar una búsqueda alfabética de wikis si el usuario está interesado en consultar información sobre los usuarios más activos.
2. **Feedback visual:** Igual de importante es que el sistema aporte algún tipo de feedback para cada acción que realiza el usuario, de forma que este sienta que la aplicación está respondiendo y se pueda interactuar con fluidez. Cuando intentamos acceder en nuestra aplicación a una sección que requiere cierto tiempo de espera (unos 5 segundos aproximadamente) debido al volumen de información que necesita cargar, aparecerá un icono tipo *loading* que dé a entender que el proceso se está ejecutando correctamente.

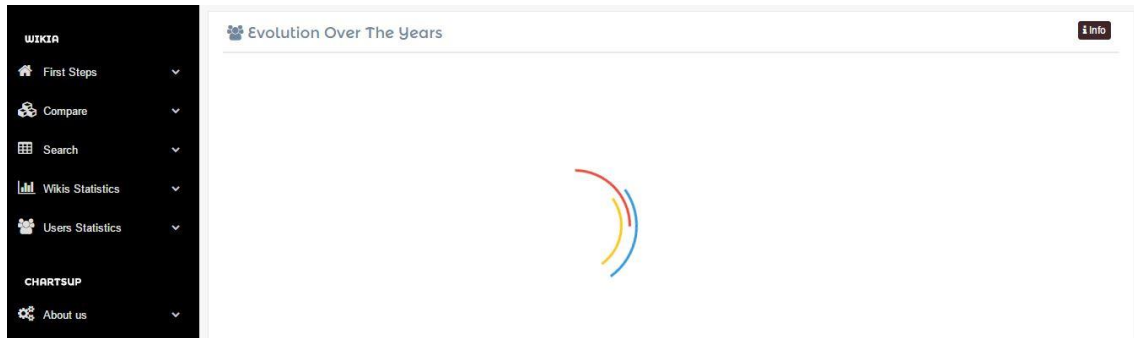


FIGURA 30. ESTADO LOADING DE PÁGINAS DE CHARTSUP

**Principio de libertad y control del usuario:** En última instancia, el usuario debe sentir que controla la aplicación. Es decir, debe poder desenvolverse en ella sin que resulte un esfuerzo y sin temor a “*romper algo*”. Por ello nuestra herramienta es sencilla, manteniendo la estructura a lo largo de todas las páginas y sin interrumpir el flujo con mensajes de aviso o confirmaciones innecesarias.

## 5.3. DISEÑO DE PANELES

A parte del diseño de la interfaz general de nuestra herramienta, es decir, menús, componentes y el resto de elementos, el punto más importante del diseño aparecía a la hora de definir los paneles. La clave de que ChartsUp cumpliera con su objetivo (presentar los datos de manera comprensible) se basa en el tipo de panel y gráficos elegidos para contar de manera precisa y sin dar lugar a error la información. Esto quiere decir que era necesario pensar detalladamente qué gráfico se iba a adaptar mejor a esa información concreta. Para ilustrar las decisiones tomadas en esta línea, vamos a explicar en las siguientes secciones los tipos de paneles de nuestra herramienta y el porqué de su uso.

### 5.3.1. INTRODUCCIÓN

Un buen gráfico muestra hechos de manera precisa, así como tendencias o cambios que afecten a la información recogida. Igual de importante es que se entienda lo que se está viendo. Un gráfico con un montón de etiquetas e información dificultan su comprensión, en vez de facilitarla.



FIGURA 31. EJEMPLO DE MAL USO DE UN GRÁFICO DE BARRAS

En el gráfico anterior [64], se observa que la meta del 31 de Marzo de 7066000 es un 17.8% mayor que 6000000, y sin embargo **la segunda barra es casi tres veces la longitud de la primera**. Esta información se presenta de manera confusa [65]. Uno podría argumentar que la atención se centra en la diferencia entre los valores en vez de en los dos valores en sí mismos. Aun así, un gráfico de barras es una elección poco acertada. Una serie temporal, que muestre un acumulado mensual, ilustraría mejor este comportamiento.

Cuando mostramos datos en gráficos se codifican en determinadas formas, colores y geometrías. Para que esto cumpla su función, sin embargo, el lector debe ser capaz de decodificar de nuevo los valores al verlo. Esta es la importancia de, por ejemplo, etiquetar los ejes [66].

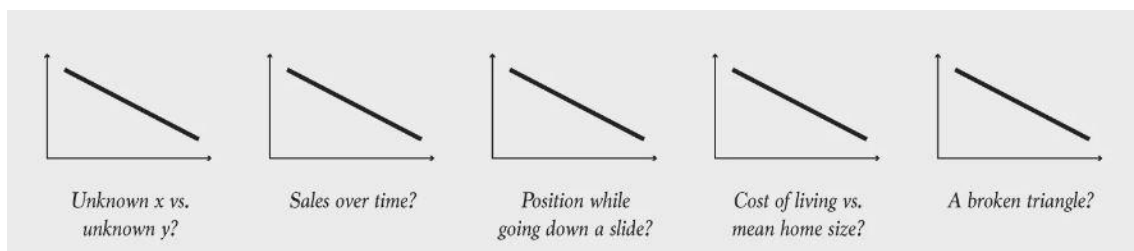


FIGURA 32. EJEMPLO DE GRÁFICOS SIN DEFINICIÓN EN SUS EJES X E Y

Es decir, hay que **explicar las escalas** para que se entienda lo que se está viendo. Sin etiquetas o explicación alguna, solo es decoración para todo aquel que no haya diseñado el gráfico. Lo mismo ocurre con las unidades. Si simplemente mostramos un gráfico con un número desnudo, el lector no sabe si se trata de unidades, porcentajes, un volumen, metros cuadrados...

Los gráficos de barras se basan en la longitud para mostrar información. Barras cortas representan valores pequeños, y barras más largas, mayores. Por este motivo es recomendable que los ejes comiencen en el valor 0. Si modificamos esa referencia, se distorsiona la realidad.

En cuanto a otros aspectos, como por ejemplo, el uso de colores, son útiles para llamar la atención sobre ciertos aspectos de la información. Pero de nuevo es importante saber elegir la gama de colores. Una gama adecuada puede ayudar a acentuar un patrón o resaltar el tema deseado.

Hay que tener cuidado al utilizar sombras, ya que si bien ayudan a resaltar diferencias, pueden dificultar la lectura de textos (pueden parecer borrosos). También es importante elegir un tipo de letra adecuado. Los espacios en blanco (zonas del gráfico vacías de texto y otros elementos) son muy importantes, ya que el cerebro humano utiliza los contrastes para distinguir unos objetos de otros. Los espacios en blanco son una de las herramientas de diseño más sencillas y elegantes para crear un contraste y facilitar así la comprensión general de un gráfico.

En conclusión, un gráfico debe contener la información necesaria como para entenderse por sí mismo.

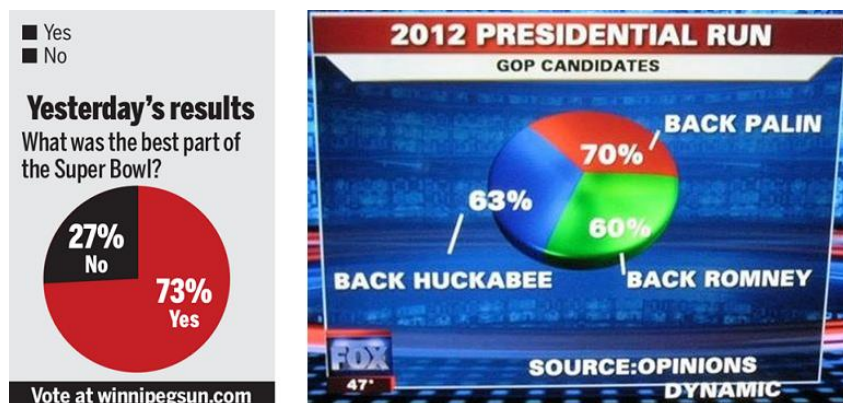


FIGURA 33. EJEMPLO DE GRÁFICOS TIPO TARTA MAL DISEÑADOS

### 5.3.2. PANELES DE CHARTSUP

A continuación mostraremos los distintos tipos de paneles utilizados en nuestra herramienta, justificando su elección. No se explicará en esta sección las listas incluidas en los paneles, sino únicamente los gráficos. La descripción sigue el orden de las opciones del menú lateral de la aplicación.

Como patrón general seguido en el diseño de estos gráficos, los fondos de todos ellos son blancos, para resaltar la información, y todos disponen del icono de guardado, *refresh*, *dataView* e info situados en la esquina superior o inferior (dependiendo del tipo de gráfico) derecha del panel, de forma que no interfieran con el gráfico en sí mismo.



Además, a la hora de representar información sobre proporciones en toda nuestra aplicación, decidimos utilizar los anillos en vez de los gráficos tipo tarta por el siguiente motivo: los gráficos tipo tarta se centran en los **tamaños relativos de las porciones entre sí**, sin dar indicaciones sobre los cambios en el conjunto. Un gráfico de anillos aborda parcialmente este problema, ya que obliga al lector a centrarse en la longitud de los arcos, más fácil de comparar que los tamaños entre las porciones.

### 5.3.2.1. Comparador de wikis

Todos los gráficos de esta página siguen la misma gama de colores, y cada una de las wikis se identifica con un color.

#### 1. Páginas por wiki

Al comparar varias wikis el primer gráfico que encontramos es el que ilustra el número de páginas totales de cada una de esas wikis. De nuevo el tipo de gráfico es de barras, definiendo en el eje X las wikis comparadas (máximo 4) y en el Y el número de páginas. De nuevo, este nos pareció el más adecuado para **comparar magnitudes**.

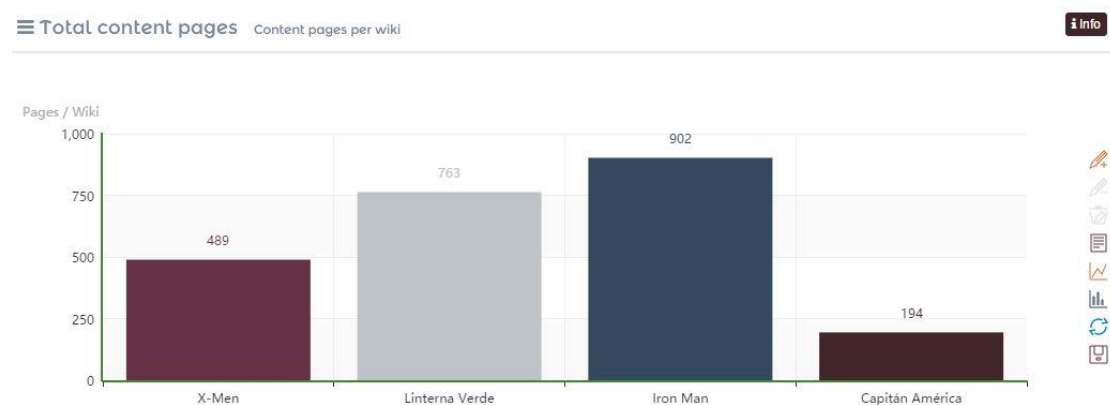


FIGURA 34. COMPARADOR DE PÁGINAS POR WIKIS

Al contener solo una serie de datos (las wikis) las opciones de visualización eran reducidas.

#### 2. Usuarios registrados, Ediciones, Media de ediciones y Ficheros subidos

El primer panel muestra el número de usuarios registrados por cada una de las wikis. En el eje X se disponen, como en estos cuatro paneles y el anterior, las wikis que se están comparando, y en el Y el valor de la categoría en su escala. En este caso, número de usuarios registrados.





FIGURA 35. USUARIOS REGISTRADOS Y EDICIONES DEL COMPARADOR DE WIKIS

En los siguientes tres paneles cambia la categoría: número de ediciones, media de ediciones y ficheros subidos. Las wikis siempre se mantienen en el eje X y en el mismo orden en los 5 gráficos para facilitar al lector la comparación.

### 3. Usuarios

En este panel se muestran el número de usuarios de los tipos principales (activos, administradores, reversores y burócratas) para cada una de las wikis comparadas. Aunque son de la misma categoría, en este caso utilizamos un **gráfico de barras agrupado, por contar con varias series de datos** (tipos de usuarios). Así, cada una se representa mediante una barra del mismo color en el eje X, y la cantidad en el eje Y.

Decidimos que un **gráfico de barras apilado** era la visualización alternativa perfecta, ya que también contiene varias series. En este caso, los segmentos de las barras se dividen por colores, y cada uno representa un tipo de usuario.

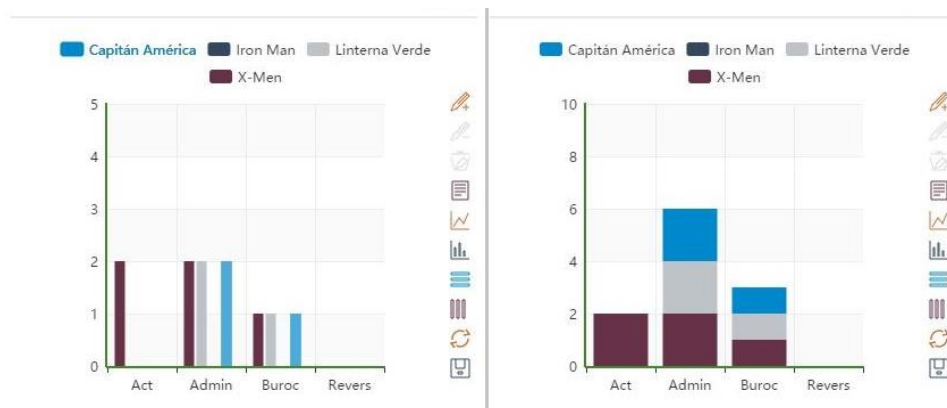


FIGURA 36. TIPOS DE VISUALIZACIÓN PARA EL TIPO DE USUARIO DE CADA WIKI COMPARADA

### 4. Actividad de usuarios

El último panel de la página muestra en qué proporción participan, en cada una de las wikis comparadas, los usuarios activos, es decir, los que han realizado algún aporte en

los últimos 30 días. Esto se representa mediante un gráfico de anillo. Este tipo de gráfico es el mejor para representar la **relación porcentual entre las partes con relación a su conjunto**.

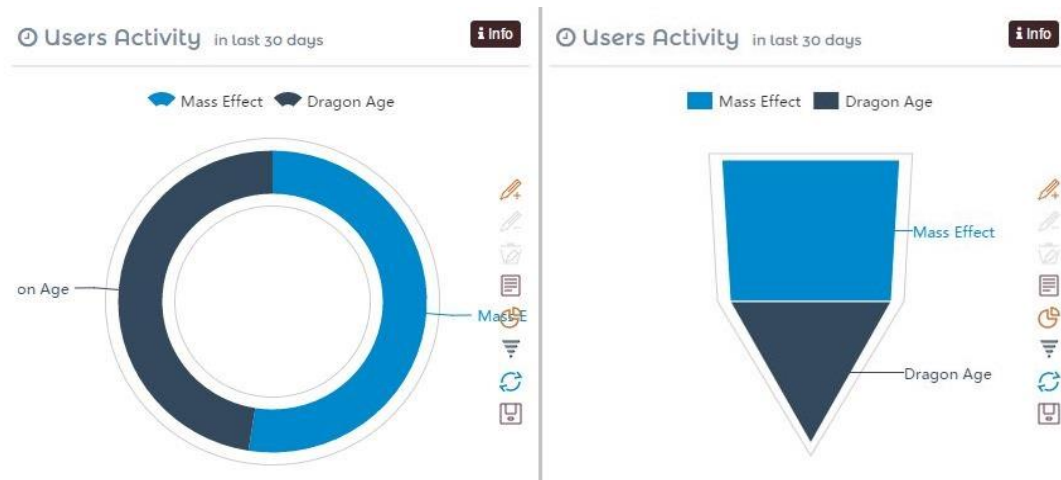


FIGURA 37. VISUALIZACIONES PARA LA ACTIVIDAD DE USUARIOS DEL COMPARADOR

El círculo representa la totalidad que se quiere observar (en el ejemplo, total de usuarios activos) y cada sección representa la proporción de cada categoría de la variable (en el ejemplo, las wikis) respecto al total. Suele expresarse en porcentajes.

La visualización alternativa es un gráfico de embudo, que representa los datos en forma de porciones del 100% y no utiliza ningún eje.

### 5.3.2.2. Perfil de una wiki

Dentro del perfil de una wiki, encontramos en primer lugar, un panel para el top 10 usuarios que más han participado en esa wiki. Este gráfico circular muestra los usuarios con sus porcentajes de participación en la wiki. Incluimos además una categoría de *Other users*.



FIGURA 38. PARTICIPACIÓN EN EL ÚLTIMO MES PARA LA WIKI

Las siguientes cuatro gráficas son gráficos de barras, ya que en todas comparamos magnitudes de varias categorías.

La estadística de páginas muestra la comparación entre el número de páginas de esa wiki, las 50 wikis con más páginas y todas las wikis registradas. La estadística de ediciones y de media de ediciones, así como de ficheros, siguen el mismo patrón: comparan las ediciones y la media de ediciones de la wiki con el top 50 y el conjunto de las wikis.

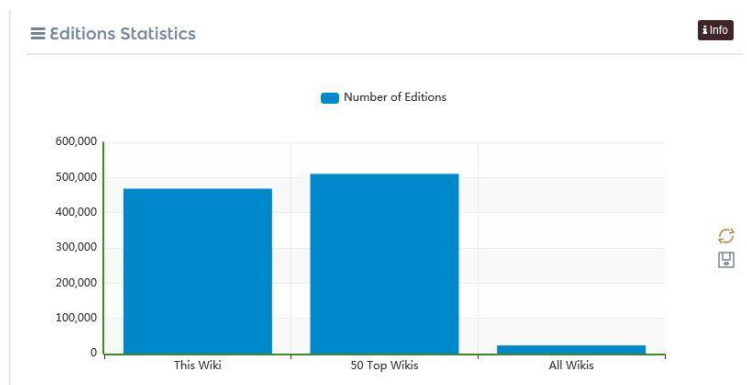


FIGURA 39. ESTADÍSTICAS DE UNA WIKI PARA LA CATEGORÍA EDICIONES

### 5.3.2.3. Estadísticas de wikis

La sección de Wikis Statistics cuenta con 9 enlaces. Los 8 primeros repiten estructura, variando solo en la categoría que se analiza (tipo de usuario, ediciones, páginas...), por lo que sólo detallaremos los paneles de una de ellas. Vamos a utilizar como categoría para el ejemplo los usuarios activos.

#### 1. Proporción de wikis con más usuarios activos

Representa con un gráfico de anillo las wikis en las que más usuarios activos hay, expresando su porcentaje.

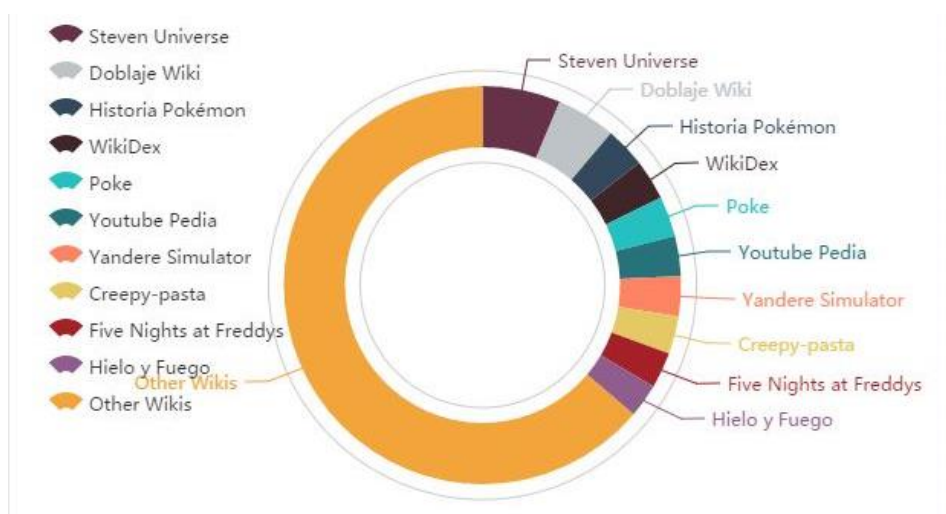


FIGURA 40. GRÁFICO DE ANILLO DE LA SECCIÓN TOP ACTIVE USERS

## 2. Top 10 wikis con usuarios más activos

Es la alternativa visual al anillo situado en la parte superior, representando las wikis en un gráfico de barras, ordenado de mayor a menor usuarios activos.

El último enlace de la lista de Wikis Statistics es distinto a los anteriores: Customizable Scatter, un gráfico de dispersión que muestra en un eje cartesiano la relación que existe entre dos variables.

Este gráfico nos informa del **grado de correlación entre las dos variables**, es decir, nos muestra si el incremento o disminución de los valores de una de las variables altera de alguna manera los valores de la otra, denominada variable dependiente y que representa generalmente en el eje vertical. El tipo de correlación se puede deducir según la forma de la nube de puntos:

- Correlación nula: No existe ninguna relación entre las variables. Se dice que ambas son independientes.
- Correlación lineal: Existe una relación lineal negativa si al aumentar los valores de la variable independiente disminuyen los valores de la variable dependiente y relación lineal positiva si al aumentar los valores de la variable independiente aumentan los valores de la variable dependiente.
- Correlación no lineal: Existe una relación entre las variables pero no es lineal.

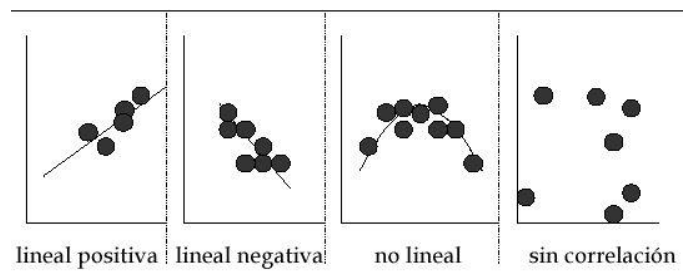


FIGURA 41. TIPOS DE CORRELACIÓN

Dado que el usuario puede elegir las variables para cada eje, esta opción era la más indicada.

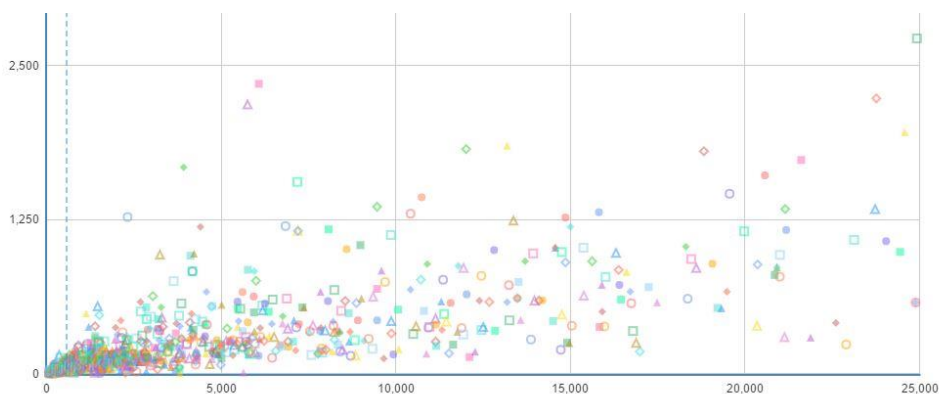


FIGURA 42. SCATTER GRAPH PARA EDICIONES (EJE X) Y PÁGINAS (EJE Y)

### 5.3.2.4. Estadísticas de usuarios

#### 1. Evolución a lo largo de los años del número de wikis en las que participan los usuarios

Dentro de las estadísticas globales de usuarios de Wikia encontramos, en primer lugar, un gráfico que representa la evolución a lo largo de los años del número de wikis en las que participan los usuarios, desde la fecha en la que empezaron a editar contenido.

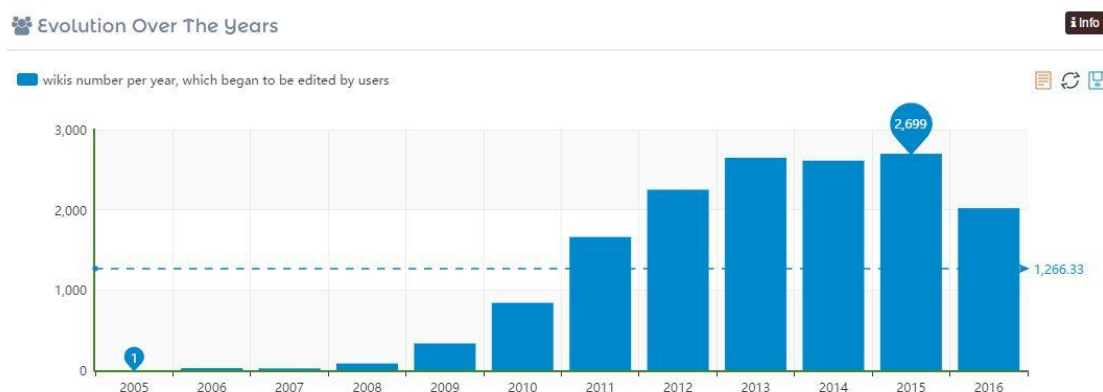


FIGURA 43. EVOLUCIÓN DEL NÚMERO DE WIKIS EDITADAS DESDE 2005 HASTA 2016

Este gráfico muestra, por tanto, una serie temporal. Elegimos este tipo de representación porque es la que **mejor refleja patrones y comportamiento de datos en el transcurso del tiempo**. En nuestro caso, el eje X dispone los años (partiendo de la primera fecha de ediciones de los usuarios) y el eje Y el número de wikis, es decir, la cantidad total de wikis editadas. El color es uniforme por no querer resaltar ningún año en particular, ya que el mínimo y máximo resultan más visibles gracias a una de las funcionalidades de la librería utilizada (ver sección 4.6). El *grid* es muy tenue, con el objetivo de no colapsar el panel con líneas innecesarias. Como particularidad, se muestra además la media.

#### 2. Usuarios más activos por número de ediciones

Como indicamos en la Introducción a esta sección, los gráficos de barras se basan en la longitud para mostrar información. Como el objetivo aquí es **mostrar una clasificación**, nos pareció que utilizar las barras ilustraba de forma precisa la cantidad de ediciones de un usuario respecto a otro, ordenados de mayor a menor participación.

En el eje X de este panel se disponen los usuarios que más han participado, mientras que en el eje Y se especifica el número de total de ediciones.

Al tratarse de un ranking era necesario distinguir cada uno de los usuarios para hacer ver que cada uno es independiente del otro. De ahí que los colores sean distintos. Dado que, el número de ediciones es muy elevado dentro de Wikia (como se puede observar por la escala del eje Y), creímos adecuado definir la cantidad exacta de ediciones del

usuario sobre cada una de las barras. De esta manera el lector se hace rápidamente una idea de qué cantidades se manejan.

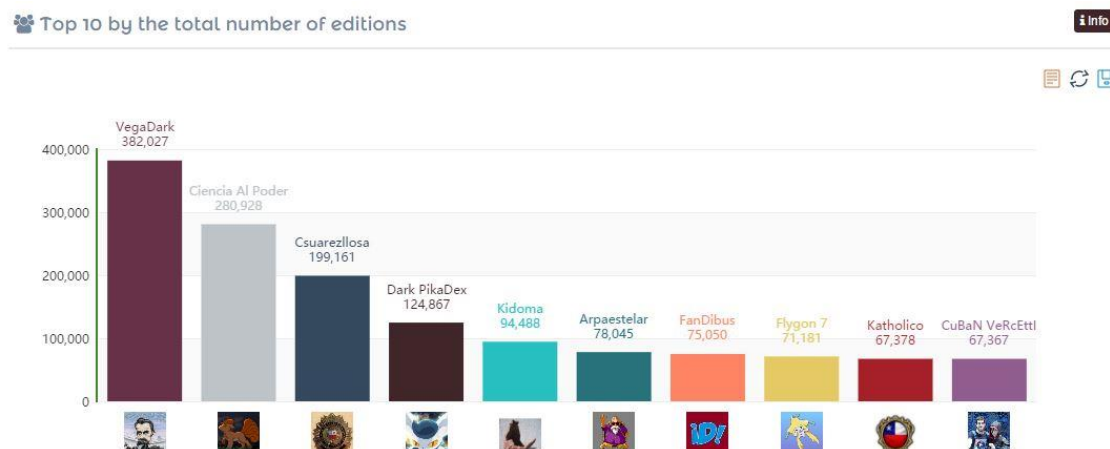


FIGURA 44. TOP USUARIOS MÁS ACTIVOS POR NÚMERO DE EDICIONES

### 3. Tipo de usuario en función de sus ediciones

El siguiente panel muestra los distintos tipos de usuarios según su **número de ediciones** en Wikia y el porcentaje de cada uno de esos tipos. La figura siguiente incluye las dos posibles visualizaciones del gráfico.

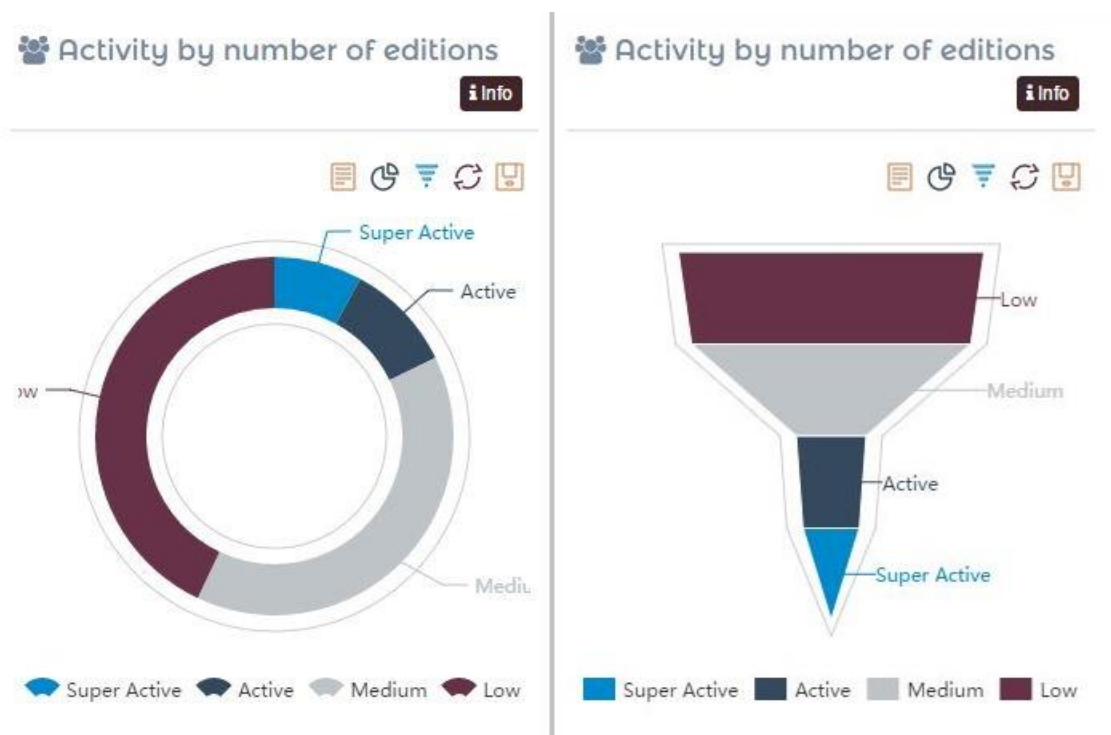


FIGURA 45. ACTIVIDAD DE LOS USUARIOS (PROPORCIÓN) POR NÚMERO DE EDICIONES

Los gráficos de anillos muestran los datos como una **proporción del total**.

Este tipo de gráfico es perfecto para **realizar comparaciones entre grupos** (actividad muy alta, alta, media, baja). El círculo está segmentado en divisiones, cada una de las cuales representa una categoría de datos diferenciada por un color. Al comparar y contrastar el tamaño de las divisiones, podemos evaluar la magnitud relativa de cada categoría. Además de mostrar la cantidad de usuarios de cada tipo, mostramos su porcentaje (pasando el cursor por encima de cada uno de los sectores).

No puedo deducir de manera sencilla si 6628 usuarios con un perfil de actividad bajo representan mucho o poco respecto del total. Sin embargo, si digo que suponen un 42.93% sí me puedo hacer una idea rápida de cuántos usuarios responden a esas características.

La alternativa que se ofrece (el gráfico de embudo) es similar en cuanto a uso al gráfico de anillo, salvo que el primero ordena las categorías de mayor a menor.

#### 4. Usuarios más activos por número de wikis en las que participan

La siguiente gráfica muestra los 10 usuarios que participan en más wikis. La decisión de utilizar un gráfico de barras con el mismo diseño que el seguido en el Panel 2 se basa en que también representamos un **ranking**. La estructura es similar: misma gama de colores para identificar a los usuarios y orden de mayor a menor en función del número de wikis en las que participan. El eje X dispone esos 10 usuarios, y el eje Y el número de wikis.

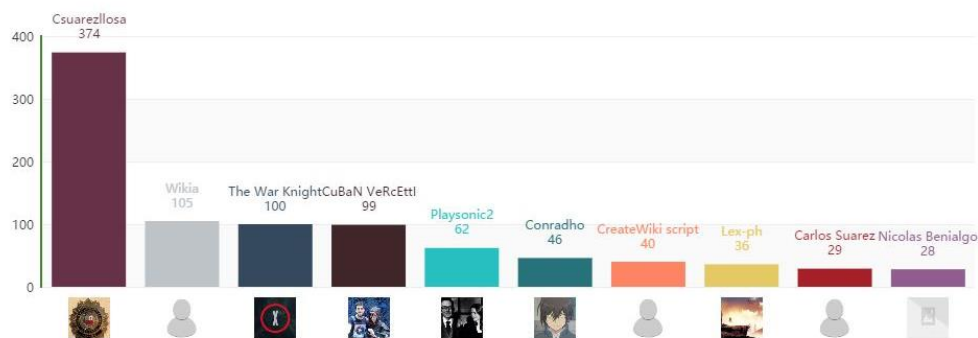


FIGURA 46. TOP USUARIOS MÁS ACTIVOS POR NÚMERO DE WIKIS EN LAS QUE PARTICIPAN

#### 5. Tipo de usuario en función del número de wikis

El último panel que encontramos en esta sección también clasifica los usuarios en función de su actividad, pero esta vez relativa al número de wikis en las que participan y no en las ediciones. Este gráfico es también tipo anillo con opción de cambiar la visualización a embudo, ya que creemos que estas dos formas son las más útiles cuando las categorías son pocas.



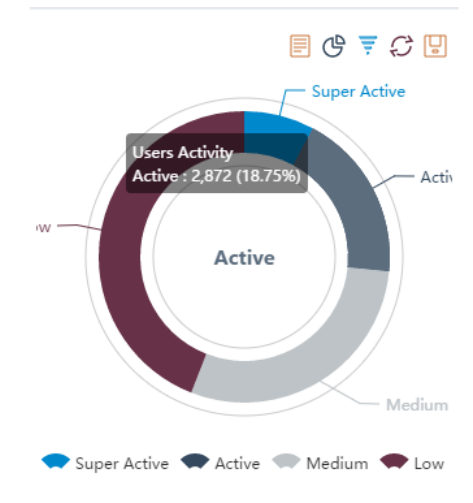


FIGURA 47. ACTIVIDAD DE LOS USUARIOS (PROPORCIÓN) POR NÚMERO DE WIKIS

### 5.3.2.5. Perfil de usuario

Si accedemos a cualquiera de los usuarios de Wikia nos encontramos con los siguientes elementos:

#### 1. Número de ediciones del usuario en cada una de las wikis en las que participa

El primero es un gráfico de barras que representa el número de ediciones de cada una de las wikis en las que participa, ordenado de mayor a menor. De nuevo las barras son la opción más precisa para ilustrar esta información, ya que muestran **datos que se clasifican en categorías nominales u ordinales**.

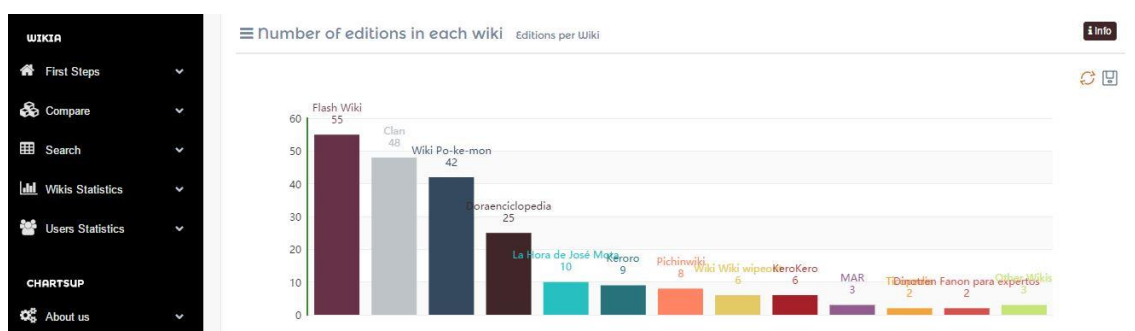


FIGURA 48. NÚMERO DE EDICIONES EN CADA WIKI EN LA QUE PARTICIPA EL USUARIO

En el eje de las X se disponen las wikis en las que participa el usuario, y en el Y el número de ediciones. Cada una de las barras tiene un color para resaltar la diferencia entre ellas. Además, sobre cada columna se detalla el número de ediciones.



## 2. Wikis en las que ha participado a lo largo de los años

Para observar el progreso a lo largo de los años de la participación de un usuario en Wikia se optó por un gráfico lineal. Este tipo de gráfico es una representación en un eje cartesiano de la relación que existe entre dos variables, **reflejando con claridad los cambios** producidos. En cada eje se representa cada una de las variables cuya relación se quiere observar.

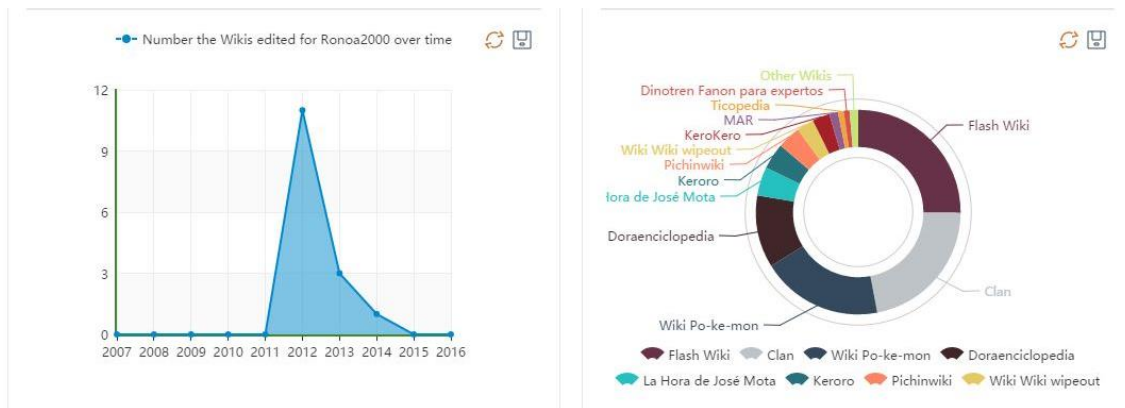


FIGURA 49. PARTICIPACIÓN DE UN USUARIO, POR AÑO (IZQ) Y EN PROPORCIÓN (DCHA)

En el eje horizontal se pone la variable que indica las unidades de tiempo (en este caso los años) y en el vertical se introduce la escala de la variable cuya variación en el tiempo queremos ver (el número de wikis en las que ha participado)

## 3. Participación en las wikis (proporción)

Por último, situado a la derecha de la serie temporal, está el gráfico de tipo anillo que representa la participación del usuario en las distintas wikis, estableciendo su porcentaje. Los gráficos de tipo anillo o tarta, como ya hemos indicado anteriormente, son los más intuitivos para mostrar proporciones.



# CAPÍTULO 6

## ARQUITECTURA DEL PROYECTO

En este capítulo se estudiarán con detalle las características más técnicas relacionadas con la arquitectura de la aplicación, desde la obtención de la información hasta la arquitectura de la base de datos.

### 6.1. RECOPILAR INFORMACIÓN

Desde el principio, uno de los problemas a los que nos enfrentamos a la hora de llevar a cabo nuestro proyecto fue el de conseguir información de la comunidad colaborativa que habíamos elegido. El problema, no obstante, puede extrapolarse a cualquiera de ellas. Hablamos de plataformas que manejan un gran volumen de contenido. Al centrarnos únicamente en Wikia en español reducíamos de entrada esa gran cantidad de datos que aún no sabíamos cómo registrar ni almacenar. Esta opción resultó la más viable ya que Wikia dispone de un listado con todas las comunidades que contiene, característica que nos permitiría tener una visión real del tiempo disponible a la hora de afrontar este proyecto.

A la hora de obtener información de las wikis encontramos dos formas que se presentan a continuación.

#### 6.1.1. WIKIA CONTENT API

Una API (*‘Application Programming Interface’*) es un conjunto de reglas y especificaciones (código) que las aplicaciones pueden seguir para comunicarse entre ellas. Las API son útiles y valiosas, ante todo, porque permiten hacer uso de funciones ya existentes en otro software, reutilizando así código que se sabe que está probado y que funciona correctamente.

Cada wiki dispone de su propia API a la que se puede acceder a través de la URL: `{wikidomain}/api/v1/`. Esto quiere decir que Wikia como tal (el conjunto de wikis) también tiene su propia API, y esta dispone de una serie de métodos (como acceso a la información sobre el ranking WAM) que no están disponible en las demás.

Vamos a explicar brevemente algunos de los términos que se manejan en la API:

- Page. Subpágina dentro de wiki. Puede ser Article, File, Category...

- Article. Subpágina dentro de wiki, que pertenece al dominio del artículo.
- File. Fundamentalmente videos e imágenes.
- Category. Corresponde a las categorías definidas por los usuarios.
- Hub. Incluye las 3 categorías en las que se agrupa Wikia: videojuegos, entretenimiento, estilo de vida
- Vertical. Es el identificador numérico (del 1 al 20) que representa una categoría (1 humor, 2 gaming, 3 entertainment...)

La obtención de información realizando llamadas a la API dependía de qué tipo de información buscásemos, ya que los métodos no son iguales entre la API “genérica” y el resto de wikis.

API que muestra datos genéricos del conjunto de todas las páginas [67]:

- Activity: información sobre el última actividad de usuarios
- Articles: contenido simplificado de los artículos
- Mercury: datos de la wiki (datos de navegación o valores *key*)
- Navigation: links de navegación
- RelatedPages: páginas relacionadas con el ID de un artículo
- Search: resultados de búsqueda sobre esa wiki
- User: detalles sobre los usuarios seleccionados
- WAM: puntuación del WAM
- Hubs: lista de hubs
- Wikis: wikis cuyo nombre o tema coinciden con una palabra clave

Si consultamos la lista de wikis con sus puntuaciones en el WAM (mostrando solo uno de los resultados):

```
{
  "wam_index": {
    "832318": {
      "wiki_id": "832318",
      "wam": "0.0000",
      "wam_rank": "5000",
      "hub_wam_rank": "1476",
      "vertical_wam_rank": "2046",
      "peak_wam_rank": "2414",
      "peak_hub_wam_rank": "946",
      "peak_vertical_wam_rank": "1029",
      "top_1k_days": "0",
      "top_1k_weeks": "0",
      "first_peak": "2015-02-19",
      "last_peak": "2015-02-19",
      "hub_name": "Gaming",
      "vertical_id": "2",
      "title": "永恆冒險 维基",
      "url": "zh.grandchase.wikia.com",
      "wam_change": "0.0000",
      "wam_is_new": "1",
      "vertical_name": "Games"
    }
  },
  "wam_results_total": "2040",
  "wam_index_date": "1464739200"
}
```

FIGURA 50. RESULTADOS DE CONSULTA PARA LA LISTA DE WIKIS CON SUS PUNTUACIÓN EN EL WAM

Vemos que devuelve el ID de la wiki, la puntuación WAM, la posición en el ranking, el título de la wiki, su URL, el número de días y de semanas que esa wiki ha estado en el top 1000 del ranking, el Hub... Además, existen otros parámetros, como la lista de administradores de cada una de las wikis, que son opcionales, pero que si se solicitan devuelven además información sobre cada uno de los administradores.

```
}
AdminItem {
  name (string): Admin name,
  avatarUrl (string): The absolute URL of the User avatar image,
  edits (integer): Number of edits,
  userContributionsUrl (string): The absolute URL of the User contributions page,
  userPageUrl (string): The absolute URL of the User page,
  userId (string): An internal identification number for User,
  since (string): Join date
}
```

FIGURA 51. INFORMACIÓN SOBRE ADMINISTRADORES DEVUELTA POR LA API DE WIKIA

Consultando la API específica de una wiki [68] vemos que se pueden extraer datos solamente de dicha wiki:

Incluye las mismas secciones que la API “genérica”, salvo WAM, Hubs y Wikis, y además incluye *SearchSuggestions* (búsqueda de frases sugeridas dentro de la wiki) Por ejemplo, para el caso de la wiki ElderScrolls, consultando la sección de Activity y fijando a 3 el límite de resultados mostrados, obtenemos lo siguiente:

```
{
  "items": [
    {
      "article": 887559,
      "user": 25356303,
      "revisionId": 2192189,
      "timestamp": 1464880270
    },
    {
      "article": 890609,
      "user": 1600847,
      "revisionId": 2192183,
      "timestamp": 1464878152
    },
    {
      "article": 395116,
      "user": 1600847,
      "revisionId": 2192182,
      "timestamp": 1464878057
    }
  ],
  "basepath": "http://elderscrolls.wikia.com"
}
```

FIGURA 52. RESULTADOS DE LA CONSULTA PARA LA ACTIVIDAD RECIENTE DE UNA WIKI

Que devuelve una lista con el ID del artículo, el ID del usuario que realizó la acción, el ID de la revisión creado para ese evento, y el timestamp de la revisión.

#### 6.1.1.1. Problema de la latencia

Inicialmente decidimos que la manera más sencilla y completa para obtener la información que queríamos era mediante consultas a la API. Después de muchos intentos de consultas simples comprobamos que efectivamente recibíamos información concreta de manera rápida. Sin embargo, cuando complicamos las *queries* para obtener grandes volúmenes de datos o simplemente información que requería consultas anidadas, la latencia se disparaba.

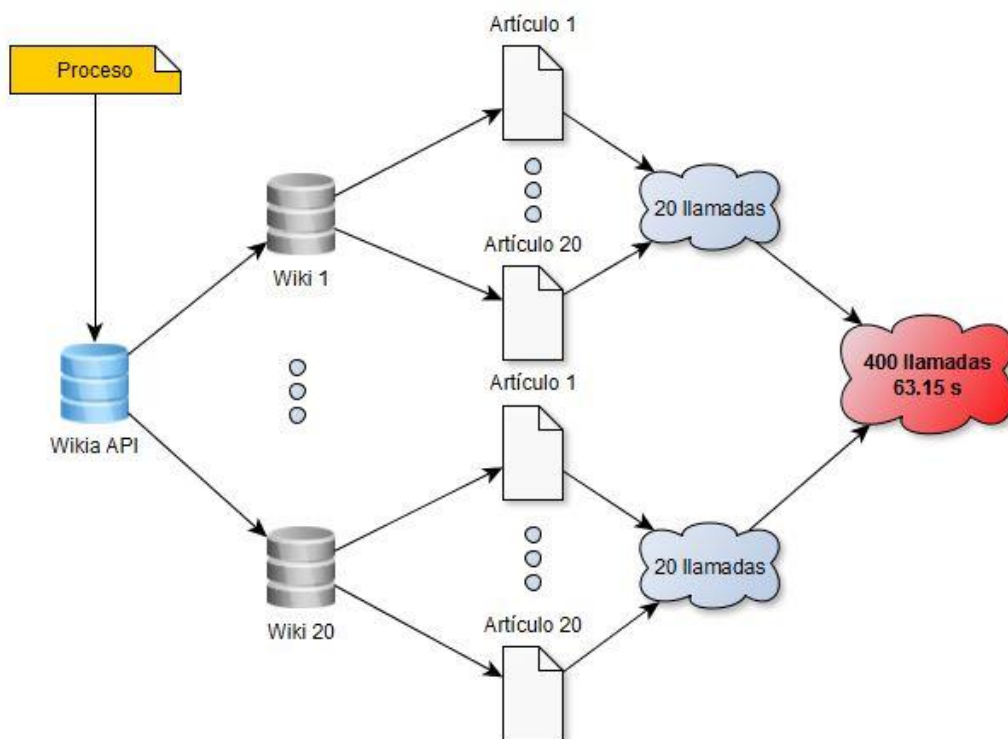


FIGURA 53. TIEMPO DE LLAMADAS A LA API DE WIKIA

El tiempo resultante de llamar a la API es alrededor de segundo y medio por cada 10 llamadas. A priori pensamos que eso no sería un problema, ya que creíamos que no haríamos más de 20 llamadas, pero al ejecutar esas sentencias más complejas nos dimos cuenta de que era un problema crítico.

Un ejemplo del problema es el siguiente: *Queremos que nuestra aplicación tenga una sección que muestre las 20 wikis más vistas en español, y que por cada una de ellas se muestren los 20 artículos más vistos.*

Para realizar esto primero debemos hacer una llamada que nos devuelve el identificador de las 20 wikis más vistas (1 llamada), por cada identificador debemos preguntar a la API de cada una por sus 20 artículos más vistos (20 llamadas), y esto sólo nos devolverá su identificador, por lo que por cada artículo debemos preguntarle otra vez a la API por los datos de ese artículo a partir de su identificador (400 llamadas). La suma de todo esto nos da 421 llamadas, lo que sería alrededor de **63,15 segundos** para recolectar toda la información.

La primera opción para intentar solventar este problema fue la de almacenar los datos obtenidos en una base de datos. Así la aplicación web muestra los datos directamente de nuestra BBDD, mientras que en un segundo plano se ejecuta un script que consiga la información mediante llamadas a la API e inserte esos datos en la BBDD. La ventaja de usar este método es que la aplicación no tiene que conectarse con la API directamente, con lo que la latencia es mucho más baja. Por otro lado, es necesario disponer de los datos previamente, y para ello tenemos que tener en nuestra BBDD una gran cantidad de información.

La realización de un script no era un problema en su desarrollo: el script se podía dejar corriendo durante días para conseguir información. Pero el nuevo problema era la gran cantidad de datos que necesitábamos. Ahora necesitamos tener toda la información posible de cualquier wiki. Unido a la gran latencia que supone la respuesta de la API tuvimos que descartar esta opción.

## 6.1.2. DATABASE DUMPS DE WIKIA

Otra forma de obtener información es mediante los database dumps disponibles desde las distintas wikis [69]. Un database dump es, generalmente, un fichero con datos que permite a un usuario tener una copia de seguridad de su información de la base de datos.

Los database dumps de Wikia son accesibles a través de la página Special:Statistics de esa wiki. Por ejemplo, para acceder a los dumps de Los Muppets, iremos a <http://muppet.wikia.com/wiki/Special:Statistics>.

Esta opción se descartó por los siguientes motivos:

1. No todas las wikis disponen de dumps. En caso de no existir hay que pedirlo a Wikia, y solo se puede solicitar uno por semana
2. Los documentos XML del dump completo de una wiki son muy pesados (varios editores de código se bloquearon al intentar abrirlos) y, aunque tiene las páginas y revisiones completas (nombres e identificadores de usuarios, comentarios) el formato en el que se presentan era engorroso (la mayor parte son los comentarios de los usuarios) y concluimos que se podía extraer la misma información de forma más sencilla usando *crawling* o parseador de URLs.



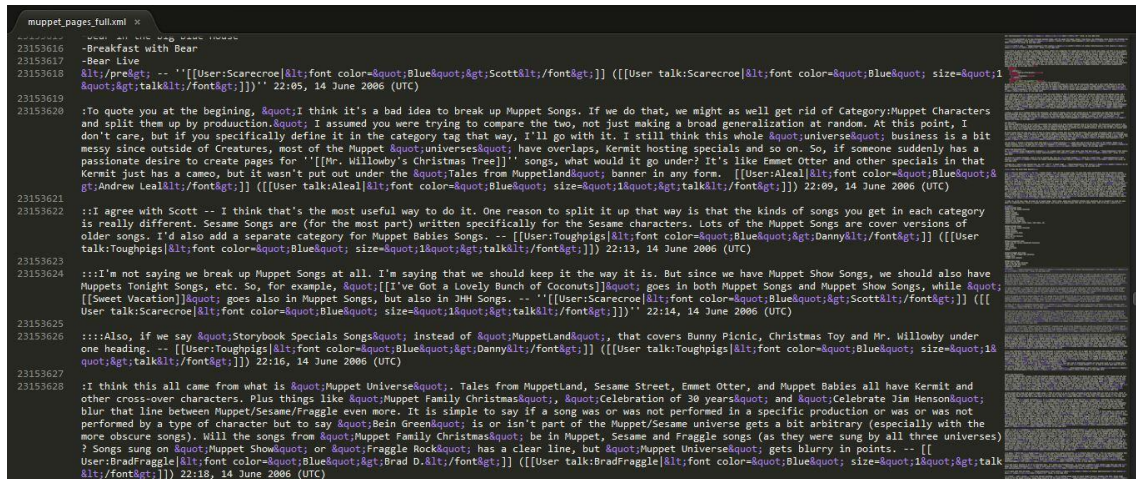


FIGURA 54. EJEMPLO DE DUMP DE LA WIKI MUPPET

### 6.1.3. PHP SIMPLE HTML DOM PARSE

La solución, por tanto, se basaba en el cambio del método de recolección de datos. En vez de mediante la API utilizando un parseador llamado PHP Simple HTML DOM Parser [70]. Este parseador **analiza una URL específica, de forma que encapsula toda la estructura HTML de cualquier página en un array**. Posteriormente, tras analizar la estructura de la página para determinar los distintos patrones, permite seleccionar la información deseada. Mediante este método podemos seleccionar toda la información que proporcionan las distintas wikis, adaptarla a nuestras necesidades e insertarla en una base de datos. La extracción es más rápida, pero como inconveniente los datos extraídos no están tan actualizados como lo son los datos que se extraen de la API, que son a tiempo real.

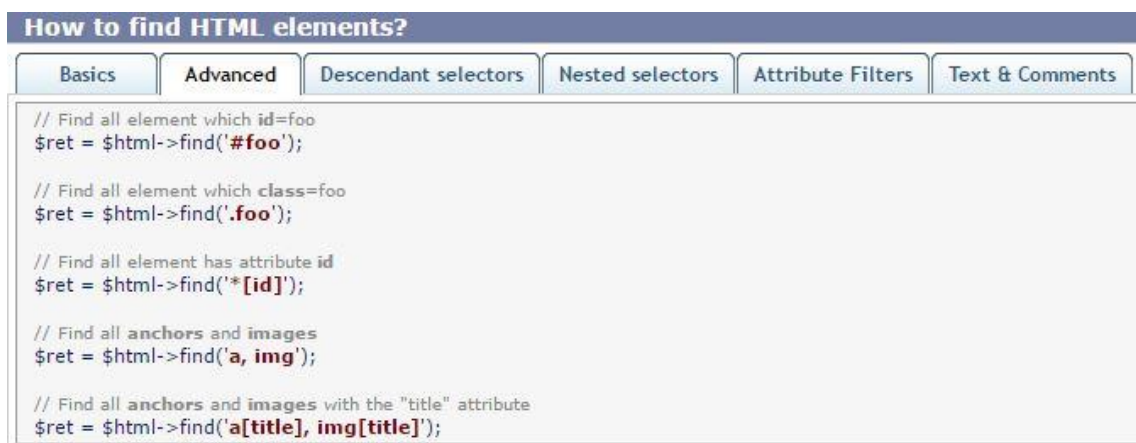


FIGURA 55. PHP SIMPLE HTML DOM PARSE DOCUMENTACIÓN



A continuación detallamos el funcionamiento del parseador, desde la obtención de los datos hasta la modificación de los mismos una vez analizada su estructura.

Si analizamos la estructura de una de la página de Wikia que contiene la lista con todas las comunidades disponibles, observamos un patrón que podíamos utilizar haciendo uso de nuestro parseador, obteniendo el nombre y las URL correspondientes a dichas wikis.

Cada wiki se encuentra dentro de un bloque `<div>` con un identificador `mw-content-text`. Dentro de este contenedor tenemos etiquetas simples `<li>` que, cuando utilizamos la función `"plaintext"`, permiten obtener el nombre de la wiki y, gracias a otra etiqueta, `<a>` con su correspondiente href, obtenemos la URL de la misma. Este sistema lo aplicamos en forma de bucle hasta que no existan más elementos.

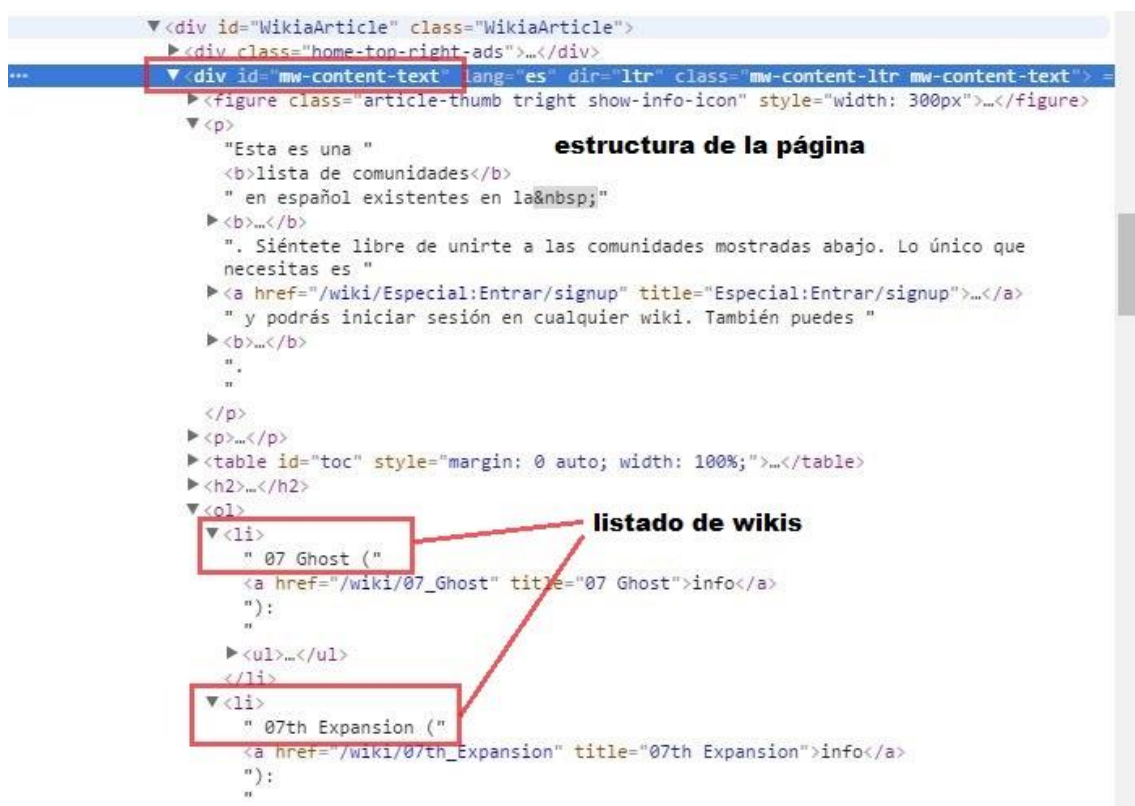


FIGURA 56. CÓDIGO DE LA PÁGINA "LISTA DE COMUNIDADES" DE WIKIA

Este ejemplo simple nos permite ver que, analizando la estructura de la página, podemos obtener los datos deseados.

Con los datos guardados en variables, ejecutamos una instrucción UPDATE o INSERT correspondiente en nuestra base de datos para tener finalmente el control de la información de forma permanente, evitando la necesidad de una conexión continua con la API de Wikia.

Al aplicar este método con las aportaciones, wikis, ediciones y todas las entidades, y comprobar que funciona correctamente y que obteníamos más información de la que nos proporcionaba la API, nos decantamos definitivamente a utilizar este método como forma para obtener información.

Tras la breve explicación de cómo funciona este parseador, pasamos a detallar los scripts que recogen esta información para nuestra aplicación web.

### 6.1.3.1. Scripts

A continuación, vamos a detallar cómo funciona el script PHP que recorre Wikia e importa la información a nuestra base de datos a modo de ejemplo, y explicaremos con detalle qué información concreta hemos obtenido haciendo crawling de Wikia.

Lo primero es encapsular la URL de la página que queremos capturar. En este caso, al tratarse de una página fija y no querer hacer recorridos en busca de otras, la definimos al comienzo. Acto seguido buscamos en el array que almacena todo el contenido los elementos que contienen el patrón “ol ul li a.extiw”, que se puede traducir de la siguiente manera: “Todo elemento con etiqueta <ol>, que dentro tenga una etiqueta <ul>, que a su vez cuente con una <li>, si dentro tiene una etiqueta <a class extiw>, entonces capturamos ese elemento”.

```
$html = file_get_html('http://comunidad.wikia.com/wiki/Listado_de_comunidades');  
  
$cadenaAuxiliar = 'Especial:Estad%C3%ADsticas';  
  
$db = new mysqli('localhost', 'root', '');  
  
$db->query("SET CHARACTER SET UTF8");  
  
if( $db->connect_errno > 0 ){  
    die('Unable to connect to database [ ' . $db->connect_error . ' ]');  
}  
  
// 2.- Elegir la base de datos  
$db->select_db('tfgdatabase');  
  
foreach($html->find('ol ul li a.extiw') as $element){  
    $url = $element->href . $cadenaAuxiliar;  
    $nombreWiki = $element->plaintext;  
    //echo $element->href . $cadenaAuxiliar . '<br>';  
  
    $query = "INSERT INTO wikis VALUES (null,'" . $nombreWiki . "', '" . $url . "', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)";  
  
    $db->query($query);  
}
```

FIGURA 57. CÓDIGO DE BÚSQUEDA DE TODAS LAS COMUNIDADES DE WIKIA

Por un lado nos quedamos con el texto plano (nombre de la wiki); Por otro, almacenamos el contenido de la etiqueta *src* para disponer de la URL que utilizaremos en un futuro para el rastreo del contenido de la misma. Finalmente, insertamos la información en nuestra base de datos y pasamos al siguiente elemento hasta llegar al final de la página.

La obtención de las aportaciones es más compleja: debemos obtener cada una de las wikis, extraer su URL, analizarla con nuestro parseador, modificar los datos para que sigan el formato adecuado (ya que manejamos tipos distintos) y realizar el resto de operaciones del caso anterior.

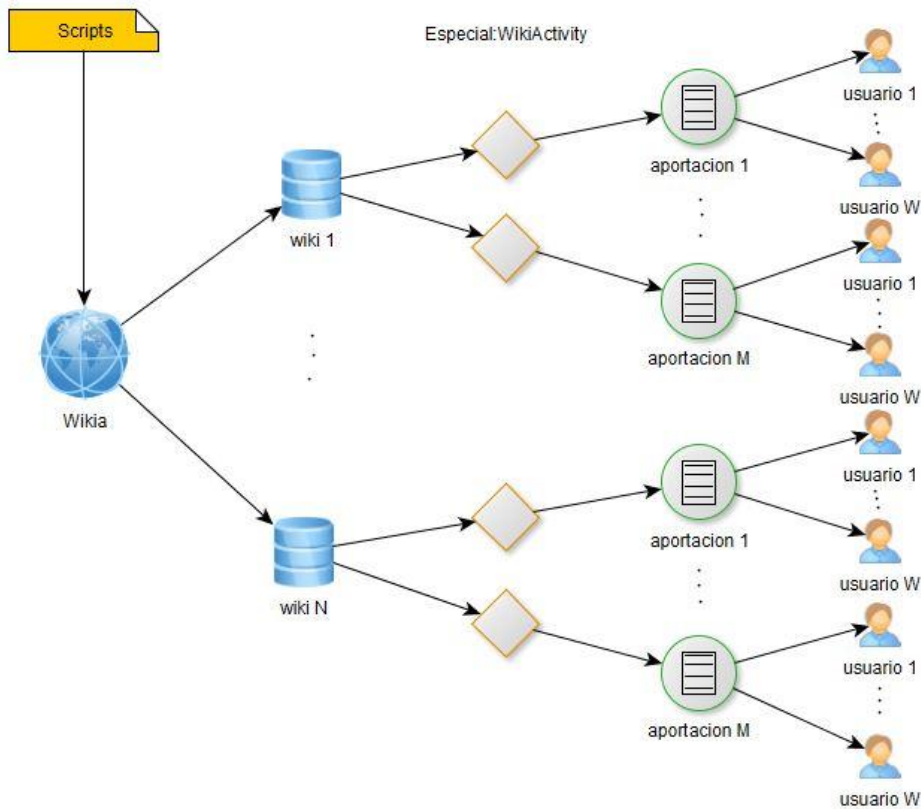


FIGURA 58. NIVELES DE ACCESO A LAS APORTACIONES DE USUARIOS PARA CADA UNA DE LAS WIKIS

Este script tarda mucho más que el de la obtención de lista de wikis (como se ve en los niveles de la figura superior): se accede a las N wikis y, por cada una de ellas, a sus M aportaciones (a través de WikiActivity). Dentro de esta página, y por cada aportación (M) accedemos a los W usuarios.



FIGURA 59. PÁGINA DE ACTIVIDAD DE LA WIKI POKEMON



FIGURA 60. FORMATO QUE SIGUE WIKIACTIVITY QUE PROPORCIONA LAS APORTACIONES DE USUARIOS



FIGURA 61. SECCIÓN DEL CÓDIGO DE PARSEO DE WIKIACTIVITY PARA OBTENER APORTACIONES

Si accedemos a cualquiera de las wikis y su página de WikiActivity, vemos que hay muchos usuarios no registrados identificados como “Editor no registrado”. Este caso se controla para no incluir a dichos usuarios.

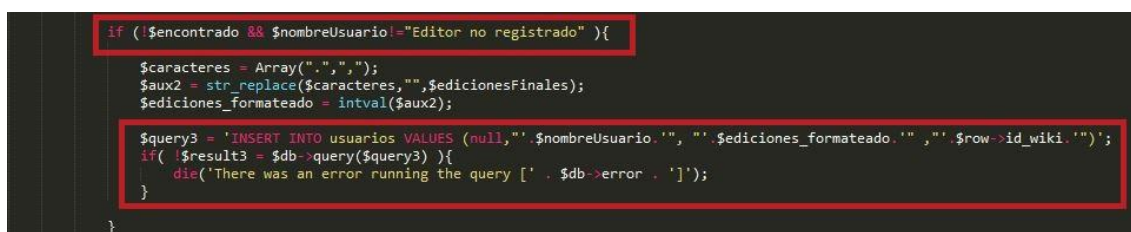


FIGURA 62. CÓDIGO PARA DESCARTAR USUARIOS NO REGISTRADOS

El resto de scripts son similares, combinando recorridos en nuestra base de datos y en las páginas de wiki con operaciones de INSERT o UPDATE. Esto es posible mediante el *include* del fichero PHP que contiene el código del parseador. Desde este fichero podemos configurar distintos parámetros, como el tamaño de la información que vamos a capturar.

No obstante, los scripts capturan información mostrada en el cliente. Pero mucha de esta información se genera después de realizar acciones por parte del usuario como, por ejemplo, pulsar un botón que realiza una búsqueda específica generando unos resultados concretos, usando o bien Ajax o bien JavaScript.

Las páginas explotadas son:

- <http://es.<nombreWiki>.wikia.com/wiki/Especial:WikiActivity>,
- <http://es.<nombreWiki>.wikia.com/wiki/Especial:Estadísticas> y
- <http://es.<nombreWiki>.wikia.com/wiki/Especial:Leaderboard>

De las cuales se ha obtenido la información siguiente:

1. Listado de **todas las wikis** de la página que lista las comunidades. Es decir, cada wiki con su logo correspondiente siempre que este sea una imagen. Si es texto plano se sustituye en nuestra aplicación por el logo de Wikia.
2. **Aportaciones que existen de los últimos 30 días de cada una de esas wikis.** Sólo recogemos las de los últimos 30 días porque son las que se presentan en la página de WikiActivity sin tener que cargar el formulario que muestra todas las demás.  
Cada una de esas aportaciones proporciona el usuario que la hizo, la dirección de su perfil, su avatar, ediciones en esa wiki y fecha en que comenzó a editarla.
3. Por cada wiki accedemos a la página Especial:Estadísticas, en la que encontramos **usuarios activos, registrados, administradores, reversores y burócratas, media de ediciones por página, ediciones desde que esa wiki fue creada, ficheros subidos y páginas totales.** El resto de campos que contienen información (verificadores y otros tipos de usuarios), si bien se podían almacenar siguiendo el mismo patrón, se descartaron porque en la mayor parte de los casos eran 0.
4. De nuevo por cada una de las wikis accedemos a su tabla de líderes, Especial:Leaderboard, siempre que dicha wiki tenga habilitada la opción de Logros. Si es así, se recorre la tabla de líderes y por cada registro se recoge el **usuario con su avatar, los puntos que lleva en esa wiki y el último logro conseguido.** Se almacena también el **avatar del logo y su descripción.** Y por cada usuario, se suman **todos los puntos** conseguidos de todas las wikis en las que ha obtenido logros.



La mayor parte de las veces ocurre que existen usuarios en la tabla de líderes que no están en las aportaciones de los últimos 30 días. Esto se soluciona buscando si existe algún usuario dentro de los líderes que no esté en aportaciones e incluyéndolo en dicha tabla.

Estadísticas de páginas	
Páginas de contenido	709
Páginas (Todas las páginas del wiki, incluyendo discusiones, redirecciones, etc.)	4.272
Ficheros subidos	1.330
Estadísticas de ediciones	
Ediciones en páginas desde que Ratchet & Clank Wiki fue instalado	15.856
Media de ediciones por página	3,71
Estadísticas de usuario	
Usuarios registrados	9.290.711
Usuarios activos (ver los miembros de este grupo) (Usuarios que han ejecutado una acción en los últimos 30 días)	4
Bots (ver los miembros de este grupo)	4
Administradores (ver los miembros de este grupo)	3
Reversores (ver los miembros de este grupo)	2
Burócratas (ver los miembros de este grupo)	2
Moderadores de contenido (ver los miembros de este grupo)	0
sotdhelper (ver los miembros de este grupo)	0
Verificadores de usuarios (ver los miembros de este grupo)	0
Oversighters (ver los miembros de este grupo)	0

FIGURA 63. EJEMPLO DE PÁGINA DE ESTADÍSTICAS DE UNA WIKI

Incluimos en los ficheros dos líneas de código que mostraran el inicio y fin del tiempo de ejecución, para calcular el tiempo total aproximado que tardan en obtener toda la información de Wikia. Las diferencias son significativas:

- Los scripts que ejecutan las inserciones iniciales en la base de datos de la lista de wikis y comparaciones tardan algo más de 10 segundos
- Las inserciones de aportaciones y logros tardan más de dos horas en ejecutarse (cada uno de los dos scripts). Esto está relacionado con los niveles de “profundidad” a los que accede el parser (ver Figura 58)
- Las actualizaciones de datos (avatar de usuarios en la tabla de líderes, título y descripción de logros, puntos de usuarios y actualización de wikis) tardan un mínimo de 30 minutos y un máximo de 6 horas.

Estos tiempos se han medido con el objetivo de automatizar los scripts, de forma que se ejecuten uno tras otro en el orden correcto y en un tiempo máximo de 10 horas.

## 6.2. BASE DE DATOS

En esta sección trataremos la estructura de la base de datos, con una descripción detallada de las tablas que la constituyen.

A la hora de llevar a cabo la implementación utilizamos un modelo relacional. MySQL (ver sección 4.11) nos sirvió como sistema de gestión de bases de datos (SGBD) y phpMyAdmin (ver sección 4.12) como interfaz de gestión de la base de datos.

Cuando empezamos a trabajar con la base de datos teníamos claro que no iba a contener un gran número de tablas, al menos trabajando con una sola plataforma colaborativa. Al incorporar otras nuevas esto podría cambiar, ya que cada una tiene una estructura específica. Intentar desde el principio diseñar una base de datos genérica implicaría una pérdida de información, por lo que se descartó.

La base de datos se divide en dos tipos de tablas: unas para guardar la información que capturamos desde Wikia (contribuciones de los usuarios, logros, wikis...) y otras para guardar la información relativa a la propia aplicación (novedades, preguntas más frecuentes...) de forma que resulte fácil acceder tanto a unas como a otras.

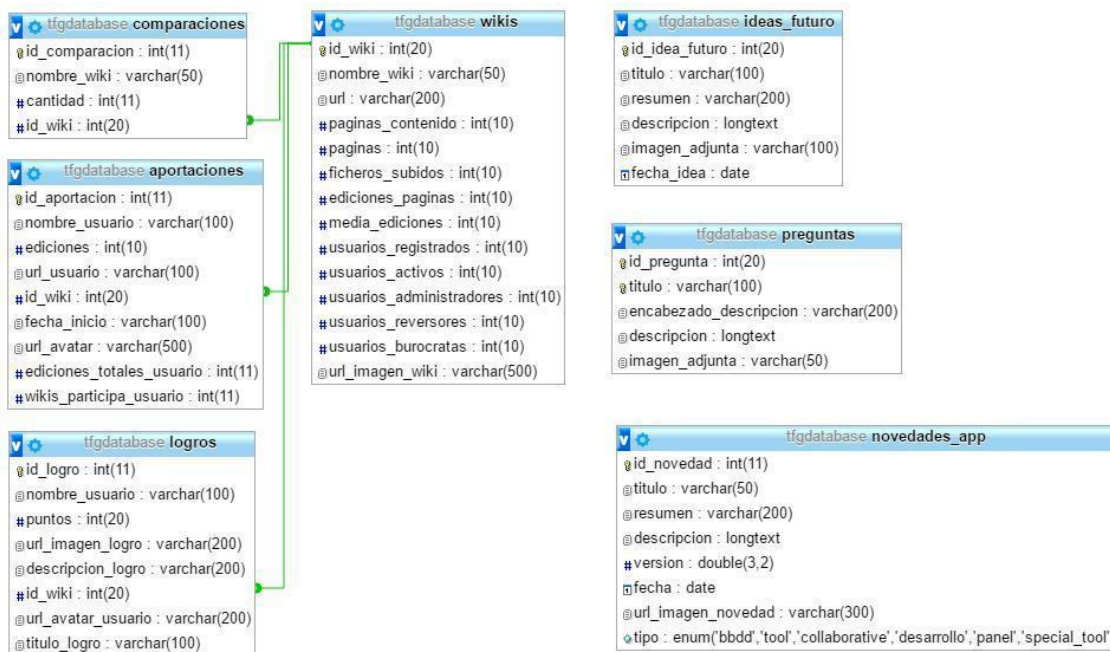


FIGURA 64. ESQUEMA DE RELACIÓN DE LA BASE DE DATOS

Todas las tablas contienen un identificador interno (*primary key*) a través del cual se accede a la información. La base de datos está formada por un total de 7 tablas, que explicaremos a continuación.

### 6.2.1. Wikis

Esta tabla almacena la información relativa a todas las wikis. Los campos se detallan a continuación:

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_wiki	Int (20)	Identificador de la wiki (interno bbdd)
Nombre_wiki	Varchar(50)	Nombre de la wiki
url	Varchar(200)	URL de acceso a la wiki
Paginas_contenido	Int(10)	Número de páginas con contenido
Paginas	Int(10)	Número de páginas en total
Ficheros_subidos	Int(10)	Número de ficheros subidos
Ediciones_paginas	Int(10)	Número de ediciones totales
Media_ediciones	Double(5.2)	La media de ediciones en la wiki
Usuarios_registrados	Int(10)	Número de usuarios registrados
Usuarios_activos	Int(10)	Número de usuarios activos
Usuarios_administradores	Int(10)	Número de usuarios administradores
Usuarios_reversores	Int(10)	Número de usuarios reversores
Usuarios_burocratas	Int(10)	Número de usuarios burócratas
url_imagen_wiki	Varchar(500)	URL de la imagen de la wiki

TABLA 2. WIKIS

Las imágenes no se almacenan directamente en la base de datos. En su lugar, guardamos la URL de la imagen, de forma que pudiéramos cargar dinámicamente en cada *img src* la imagen y así ganar en eficiencia.

El campo usuarios activos se almacena directamente como un dato cuantitativo. Sin embargo, en otras secciones de la aplicación, como cuando se visualizan los gráficos que expresan el porcentaje de actividad de los colaboradores, implantamos una serie de requisitos internos que determinan el grado de actividad en función de sus ediciones y del número de wikis en las que participa.



## 6.2.2. Aportaciones

Esta es la tabla que contiene información sobre las aportaciones por parte de los usuarios de Wikia.

Inicialmente, nuestra idea era generar una tabla específica para usuarios, otra para los aportes que realizaban y una más con las claves referenciadas. Es una manera más estructurada de tener la información y evitar, entre otras cosas, duplicados. Pero capturar información mediante el parseador, como explicamos en el punto anterior, hace que sea mucho más difícil alcanzar esa estructura. Al realizar consultas e insertar dicha información en fragmentos de código js la latencia aumenta de manera significativa.

Así, decidimos agrupar todo el contenido relativo a los usuarios en una única tabla. Duplicamos el nombre del usuario, pero consideramos que esto era un mal menor comparado con los resultados de implementar la primera opción.

La tabla sigue esta estructura:

Campo	Tipo	Descripción
Id_aportación	Int (11)	Identificador de la aportación (interno en la bbdd)
Nombre_usuario	Varchar(100)	Nombre del usuario que realiza la aportación
ediciones	Int(10)	Número de ediciones del usuario en la wiki
url_usuario	Varchar(100)	URL de la ficha del usuario
Id_wiki	Int(20)	Id que referencia al id de la wiki
Fecha_inicio	Varchar(100)	Fecha en la que comienza a editar en la wiki
url_avatar	Varchar(500)	URL de la imagen del avatar del usuario
Ediciones_totales_usuario	Int(11)	Número total de ediciones del usuario
Wikis_participa_usuario	Int(11)	Número total de wikis en las que participa

TABLA 3. APORTACIONES

### 6.2.3. Logros

Logros contiene la información de las insignias conseguidas por parte de los usuarios. Siguiendo el mismo planteamiento que en la tabla Aportaciones, la idea inicial era utilizar la tabla de usuarios para relacionarla con logros y conseguir una estructura uniforme y coherente. Es decir, que la existiera una tabla con el identificador del logro y el del usuario. Por la misma razón que ocurre en Aportaciones, descartamos esa idea.

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_logro	Int (11)	Identificador del logro interno bbdd
Nombre_usuario	Varchar(100)	Nombre del usuario que consigue el logro
puntos	Int(20)	Puntos conseguidos por el usuario
url_imagen_logro	Varchar(200)	URL de la imagen del logro
Descripción_logro	Varchar(200)	Descripción del logro
Id_wiki	Int(20)	Id que referencia al id de la wiki
url_avatar_usuario	Varchar(200)	URL imagen avatar usuario
Titulo_logro	Varchar(100)	Título del logro

TABLA 4. LOGROS

### 6.2.4. Comparaciones

Esta tabla contiene información de todas las comparaciones entre wikis. El objetivo es medir qué wikis resultan más interesantes para los usuarios, de manera que se sume uno a la cantidad de esa wiki cuando se compara con otra. Así, posteriormente, las más comparadas se incluyen en un ranking.

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_comparacion	Int (11)	Identificador de la comparación (interno bbdd)
Nombre_wiki	Varchar(50)	Nombre de la wiki
Cantidad	Int(11)	Cantidad de veces que se ha utilizado para ser comparada
Id_wiki	Int(20)	Id que referencia al id de la wiki

TABLA 5. COMPARACIONES

### 6.2.5. Novedades\_app

La información sobre progresos y notificaciones que se llevan a cabo en la aplicación web se almacena en esta tabla.

Las notificaciones pueden ser de varios tipos: algunas implican procesos en la base de datos (más wikis disponibles), mejora o aumento de la información, nuevos desarrollos, herramientas o eventos. Así conseguimos implantar un sistema interno de notificaciones que hace que el usuario pueda tener conocimiento de las novedades en las que estamos trabajando o que queremos publicar.

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_novedad	Int (11)	Identificador de la notificación (interno bbdd)
Titulo	Varchar(50)	Título de la notificación
Resumen	Varchar(200)	Resumen de la notificación
descripcion	Longtext	Descripción detallada de la notificación
Versión	Double(3,2)	Versión de la novedad en caso de una herramienta y que hubiera actualizaciones
Fecha	date	Fecha en la que se da de alta la novedad
url_imagen_novedad	Varchar(300)	URL de la imagen asociada al tipo de notificación novedosa
tipo	Enum(“bbdd”, “tool”, “collaborative”, “develop”, “panel”, “special_tool”)	Tipos de notificación: base de datos, herramienta estándar, social, desarrollo, panel nuevo y herramienta especial

TABLA 6. NOVEDADES DE LA APLICACIÓN

### 6.2.6. Preguntas

Tiene la información de las preguntas más frecuentes relacionadas con el uso de la aplicación. Este tipo de información nos permite saber qué cuestiones aparecen más confusas para los usuarios, proporcionando un soporte básico, rápido y sencillo sobre la aplicación. La tabla tiene los siguientes campos:

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_pregunta	Int (20)	Identificador de la pregunta (interno bbdd)

titulo	Varchar(100)	Título de la pregunta
Encabezado_descripcion	Varchar(200)	Encabezado de la descripción
descripcion	Descripción	Descripción o respuesta de la pregunta
Imagen_adjunta	Imagen_adjunta	URL de la Imagen adjunta

TABLA 7. PREGUNTAS

### 6.2.7. Ideas\_futuro

Almacena información de todas las ideas de desarrollo futuro sobre la aplicación.

A priori podría parecer poco coherente contar con una tabla cuyo uso es relativo al desarrollo de la herramienta, pero la información que se guarda en esta tabla se refleja en un panel específico de la aplicación web. El grado de detalle sobre esta información y cantidad mostrada es mucho mayor que el de una simple notificación. Por ejemplo, si incorporamos otras plataformas como Github o StackOverflow, se incluirían las pautas en esta sección, de tal forma que sirva como túnel conductor e informativo constante entre usuario y aplicación.

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
Id_idea_futuro	Int (20)	Identificador de la idea de futuro (interno bbdd)
titulo	Varchar(100)	Título del proyecto
resumen	Varchar(200)	Resumen del proyecto planificado
descripcion	Descripción	Descripción detallada de los planes de proyecto
imagen_adjunta	Imagen_adjunta	URL de la Imagen adjunta por si fuera necesario
fecha_idea	date	Fecha de publicación del proyecto de futuro

TABLA 8. IDEAS PARA DESARROLLOS FUTUROS

## 6.3. DISEÑO WEB

Dado que este es un proyecto sobre visualización de información, uno de los puntos más importantes de la aplicación es **mostrar esa información de una manera atractiva para el usuario**, utilizando elementos llamativos que cuenten con funcionalidades y opciones distintas a los habituales gráficos de barras.

Una de las prácticas más extendidas a día de hoy es el diseño web *responsive* o adaptativo, que busca la correcta visualización de una misma página web en distintos dispositivos. Este es el principal motivo por el que elegimos utilizar Bootstrap (sección 4.8). Esta herramienta, además, facilita la implementación de una web gracias a sus estructuras predefinidas. El uso de HTML está complementado por la utilización de CSS para personalizar el estilo y JavaScript para funcionalidad interna. Durante el análisis de librerías (ver sección 2.4.2.2) mencionamos algunas que permiten elaborar gráficos con estas características.

La variedad de gráficos de Echarts es interesante porque el gráfico elegido para mostrar un tipo de información debe adaptarse de forma que **se entienda lo que se está viendo**. Por ejemplo, no parece adecuado utilizar un gráfico tipo *donut* para explicar la evolución a lo largo del tiempo de cierto parámetro. En este caso lo más intuitivo es utilizar una serie temporal.

Sin embargo, si queremos jugar con ese gráfico para que no se trate del típico ejemplo que encontramos en cualquier página, esta librería ofrece opciones para el gráfico que se salen un poco del formato estándar, sin perder por ello su función. Para nosotros era muy importante ofrecer al usuario unos gráficos interactivos, que permitieran modificar valores en tiempo real, trazar líneas para una mejor comprensión, cambiar el *layout* o permitir la descarga en formato PNG del gráfico. Por eso la utilización de esta herramienta tan potente se adaptó perfectamente a nuestras necesidades.

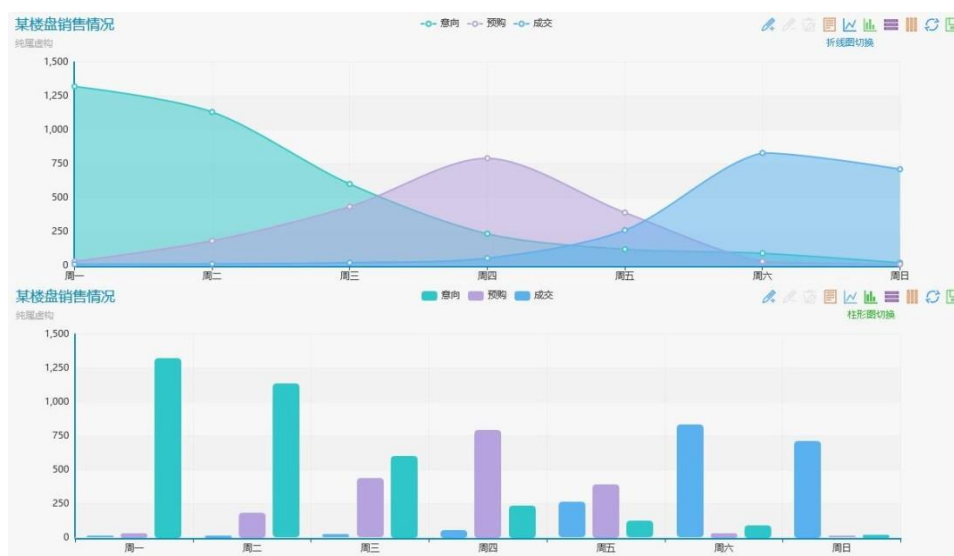


FIGURA 65. TRANSICIÓN DE UNA VISUALIZACIÓN A OTRA DE ECHARTS

Echarts también permite modificar los datos en tiempo real a través de un botón situado en la parte superior derecha de todos los gráficos. Esta característica puede resultar útil a aquellos usuarios que quisieran observar el comportamiento de una gráfica con datos libres. El botón de *refresh* siempre permite volver a su estado original.

Otro punto interesante es que permite modificar la disposición de elementos como el título, la leyenda y la suite de botones de funcionalidades. Esto nos deja espacio para distribuir todas las opciones y aprovechar al máximo el espacio. Además, el estilo encaja perfectamente con el aspecto general de la interfaz, proporcionando uniformidad a la aplicación.

El lenguaje utilizado es, como ya mencionamos anteriormente, JavaScript. Los estilos se aplican con CSS, y todo ello se incrusta en un contenedor que se basa en un canvas. La siguiente figura expone un ejemplo en el que se observa la especificación de una anchura de columna de 6/12. Dentro de esta columna se definen dos paneles, uno con un identificador *echart\_donut3* y otro *echart\_donut*.

```
<div class="col-md-6 col-sm-6 col-xs-12 profile_left">
  <div class="x_panel">
    <div class="x_title">
      <h2><i class="fa fa-bars"></i> User Participations <small>participations per year</small></h2>
    </div>
    <div class="clearfix"></div>
    <div class="x_content">
      <div id="echart_donut3" style="height:350px;"></div>
    </div>
  </div>
</div>
<div class="col-md-6 col-sm-6 col-xs-12 profile_left">
  <div class="x_panel">
    <div class="x_title">
      <h2><i class="fa fa-bars"></i> User Participation <small>percent in each wiki</small></h2>
    </div>
    <div class="clearfix"></div>
    <div class="x_content">
      <div id="echart_donut" style="height:350px;"></div>
    </div>
  </div>
</div>
```

FIGURA 66. DEFINICIÓN DE LA ESTRUCTURA DE DOS PANELES DE ECHARTS

Vamos a centrarnos en el panel que contiene un gráfico con el id='echart\_donut3' contenido en el fichero **fichaUsuario.php**. Este fragmento de código define el gráfico dentro del panel *User Participations: participations per year* de un usuario.

```
<script>
var myChart10 = echarts.init(document.getElementById('echart_donut3'), theme);
myChart10.setOption({

  tooltip : {
    trigger: 'axis'
  },
  legend: {
    // orient : 'vertical',
    // x : 'right',
    // y : 'bottom',
    data:['<?php echo'.$row->nombre_usuario.''; ?> participations']
  },

```

FIGURA 67. CÓDIGO DE DEFINICIÓN DE LAS OPCIONES DE UN PANEL (1)

Consultando la API vemos que este tipo de gráfico dispone de varios atributos:

- Tooltip: hace referencia a la “caja” que aparece cuando pasamos el cursor sobre uno de los elementos.
- Legend: nuestra leyenda, cuyo formato dependerá de la clase de elementos que estemos utilizando. En este caso solo queremos que aparezca el nombre del usuario e indicar sus participaciones.
- Toolbox: define la suite de herramientas que podemos aplicar sobre nuestro gráfico. En este caso tenemos habilitadas *saveAsImage* (guardar imagen) y *restore* (refrescar el gráfico)
- xAxis: eje X del gráfico
- yAxis: eje Y del gráfico
- Calculable: esta función nos permite recalcular automáticamente los datos cuando omitimos o movemos componentes.
- Series: la parte más importante del script, que contiene la información que queremos mostrar, datos y relaciones.

```
calculable : true,
series : [
  {
    name: '<?php echo'.$row->nombre_usuario.''; ?> participations',
    type: 'line',
    itemStyle: {normal: {areaStyle: {type: 'default'}}},
    data : [<?php echo'.$contador2007.'';?>, <?php echo'.$contador2008.'';?>,
    <?php echo'.$contador2009.'';?>, <?php echo'.$contador2010.'';?>, <?php echo'.$contador2011.'';?>,
    <?php echo'.$contador2012.'';?>, <?php echo'.$contador2013.'';?>, <?php echo'.$contador2014.'';?>,
    <?php echo'.$contador2015.'';?>, <?php echo'.$contador2016.'';?>],
  }
]

```

FIGURA 68. CÓDIGO DE DEFINICIÓN DE LAS OPCIONES DE UN PANEL (2)

Este es un ejemplo muy sencillo. Sin embargo, podemos comprobar que hay fragmentos de código PHP en el interior del script JS. Esto se debe a que los datos que utilizan todos nuestros gráficos son **dinámicos** y se obtienen mediante consultas a nuestra base de datos. En este caso concreto es un dato el que obtenemos tras la consulta, pero otros de los gráficos que utilizamos son más complejos y requieren consultas que necesitan recorridos. Para ello necesitamos establecer dentro del atributo *data* bucles que permitan

introducir, por ejemplo, las wikis en las que participa un usuario junto con sus ediciones.

De todo ello se deduce que no podemos disponer de la misma estructura para varios gráficos. Cada definición corresponde a un único panel. Si añadimos esto a que al mínimo error todos los gráficos dejan de funcionar correctamente, la depuración se convierte en una tarea que hay que llevar a cabo con mucho cuidado. El resultado, no obstante, es bueno: gráficos poco habituales que permiten al usuario jugar mediante la interacción, y con un tiempo de carga reducido.

## 6.4. SERVIDOR WEB

En nuestro caso, como ya comentamos en el capítulo 4, utilizamos XAMPP para llevar a cabo el desarrollo del proyecto la mayor parte del año. Esto quiere decir que para probar nuestro código simplemente teníamos que incluir los ficheros dentro de la ruta especificada en la instalación del XAMPP, arrancar los servicios MySQL y Apache, y acceder a *localhost* para visualizar la página con la que estuviéramos trabajando.

A finales de Mayo, sin embargo, decidimos utilizar un servicio de hosting y dejar de trabajar sólo a nivel local. Elegimos Hostinger, que proporciona 2GB de espacio en disco sin ningún coste, y creamos nuestra página web bajo es subdominio <http://chartsup.esy.es/>.





# CAPÍTULO 7

## EVALUACIÓN CON USUARIOS

Esta fase incluye las evaluaciones de usuarios con la primera versión de la aplicación ya operativa. Así se comprobará la experiencia de los usuarios y se podrá ver qué objetivos se han logrado satisfactoriamente y qué mejoras deberíamos incluir.

### 7.1. OBJETIVOS DE LA EVALUACIÓN

El objetivo principal de la evaluación es poner a prueba la funcionalidad de la aplicación, no sólo a nivel de interfaz, sino también a nivel de claridad de los datos con los que trabajamos, y que constituyen la parte fundamental del proyecto. Queremos analizar cómo se manejan los usuarios con la herramienta y de qué forma utilizan las funcionalidades que ofrecemos. Para ello, después del feedback de los usuarios, daremos respuesta a las siguientes cuestiones:

- Relativas a la calidad de la información:
  - ¿Eres capaz de comprender la información que se muestra?
  - ¿Qué información que no está disponible incluirías en la aplicación?
  - ¿Qué información consideras poco útil dentro de la misma?
- Relativas a la calidad de la interfaz:
  - ¿Encuentras atractivo el aspecto general de la herramienta?
  - ¿Hay alguna tarea demasiado compleja como para llevarla a cabo?
  - ¿Es complicado localizar algún elemento concreto dentro de la interfaz?

### 7.2. CARACTERÍSTICAS DE LA EVALUACIÓN

Las evaluaciones tendrán una duración máxima de 40 minutos. En cada una de ellas deberán cumplirse los puntos detallados a continuación:

1. El moderador dará una breve explicación a los participantes del tipo de herramienta que se está desarrollando y sus principales objetivos.
2. Durante la sesión, el moderador tomará nota de todos los comentarios que irá proporcionando el usuario sobre cada una de las tareas que lleve a cabo, poniendo especial atención a los aspectos negativos o dificultades halladas.

3. Al terminar la misma, el usuario dará *feedback* general al moderador, que también quedará incluido en las notas.

La evaluación será realizada por tres usuarios cuyos perfiles detallaremos a continuación. Aunque el tipo de evaluación más útil para este tipo de proyectos es, quizá, una prueba de usabilidad con test de usuarios como la que hemos seguido durante este proceso, el número de participantes óptimo para pruebas de este estilo suele situarse en torno a 5 personas. Estas pruebas, además, pueden resultar más breves que las realizadas para esta evaluación, ya que el proceso consiste en definir unas tareas concretas que el participante tiene que llevar a cabo.

En nuestro caso, sin embargo, optamos por reducir el número de participantes pero no establecer unos ejercicios concretos, sino permitir al usuario navegar y probar la herramienta sin más objetivo que el intentar explorar todas las funcionalidades de la misma. Este sistema nos aporta mucho más *feedback* y de mayor calidad, ya que los participantes disponen libremente de la aplicación.

Los participantes no se adaptan perfectamente a los perfiles definidos en la fase de modelado de personas (ver sección 5.2.2), pero si cumplen con los requisitos fundamentales. Lo ideal para esta evaluación habría sido contar con sociólogos o analistas, personas con perfiles estrechamente relacionados con el estudio de información, más concretamente, relacionada con comunidades colaborativas online. Sin embargo, de nuevo por problemas de agenda, no pudimos incluir estos perfiles, y tuvimos que trabajar con los que se detallan a continuación:

- **Usuario 1**, Project Manager en consultoría Big Data, es el enlace entre su empresa y los clientes a los que dan servicio. Por este motivo, la forma de contar qué avances (o retrocesos) ha habido en un determinado proyecto es muy importante. Dado que trabaja diariamente con herramientas de visualización como Tableau, pensamos que su perfil podía ser interesante para analizar nuestra herramienta, centrándonos en el aspecto de la interfaz.
- **Usuario 2** es el encargado de la gestión de equipos y procesos de IT en una multinacional. Su trabajo consiste, entre otras cosas, en la optimización y automatización de procesos. Para ello tiene que analizar las incidencias, estudiar reportes y proporcionar soluciones que mejoren el rendimiento de su equipo. Esto implica trabajar con grandes cantidades de datos, utilizando herramientas tipo *dashboards* que le permitan ver rápidamente el estado de las tareas, el incremento de fallos, etc. La evaluación de este usuario se centró fundamentalmente en analizar la calidad de los datos con los que trabajamos en nuestra aplicación.
- **Usuario 3**, estudiante de Máster en Informática en Universidad Complutense de Madrid. Aunque el máster en Informática que está completando este usuario no está relacionado con el análisis de información, es usuario activo y muy frecuente de Wikia. Contribuye de manera activa en comunidades relacionadas

con videojuegos, y conoce con detalle las características concretas de esta plataforma, por lo que fue el primer candidato a la hora de evaluar nuestro software.

Todas las entrevistas se realizaron en entornos aislados y tranquilos, evitando así posibles distracciones. De esta forma, los entrevistados expusieron sus dudas y problemas en voz alta para que el moderador las resolviera y guiara durante el proceso.

En cuanto a la aplicación, dado que la evaluación no es sobre un prototipo, se utilizó la herramienta ya instalada en el ordenador del moderador. El moderador es el responsable de facilitar ayuda durante del proceso de evaluación con el usuario. No establecimos una lista de tareas para realizar con cada uno de los usuarios, sino que fue el propio usuario el que decidió qué consultas quería llevar a cabo. De esta forma obtuvimos una visión global sobre el entorno e información de la aplicación desde 3 puntos de vista distintos. Durante el proceso, el moderador sólo podía interrumpir al usuario por motivos de fuerza mayor, es decir, conexión a internet interrumpida, error en la carga de una página, etc. Tras finalizar la evaluación, el usuario dio *feedback* al moderador, que tomó nota de todos los puntos que se tuvieron en cuenta.

## 7.3. RESULTADOS DE LA EVALUACIÓN

Los 3 usuarios respondieron de forma fluida al uso de la aplicación. 2 de los 3 entrevistados estaban acostumbrados a utilizar herramientas de visualización de información de manera habitual, pero todos se adaptaron rápidamente a la navegación a través de la interfaz.

### 7.3.1. INTERFAZ GRÁFICA

#### Aspectos positivos:

Todos los usuarios coinciden en que la estética de la misma es muy interesante y atractiva. Cada gráfico se encuentra dentro de su propio panel, incluyendo la leyenda y la suite de opciones para ese gráfico, y todos coinciden en que el diseño es claro. Una de las funcionalidades que ha resultado más útil según los usuarios es la opción de poder descargar un gráfico en particular, para poder incluirlo en un estudio propio, ahorrándose el paso de, por ejemplo, hacer una captura de pantalla. Además, la interactividad de los gráficos permite transformar algunos de ellos, de manera que si alguno de ellos presenta la información de una manera que resulte confusa o no se adapta a nuestras necesidades podemos modificar su aspecto en un clic.

Así mismo, los usuarios no encontraron dificultad para acceder a los elementos del sistema y pudieron completar las tareas que se habían propuesto con éxito. Ya que dimos total libertad al usuario para “cacharrear” con la aplicación, ellos decidieron qué funciones les parecían interesantes y cuáles no.

#### Aspectos negativos:

Lo primero que destacaron 2 de los 3 usuarios es que el nombre de la aplicación no resultaba amistoso ni sencillo de recordar. El tercer entrevistado indicó que el nombre no le parecía relevante si encontraba útil la aplicación.

Otra dificultad con la que se toparon los usuarios es que hay ciertos **textos que no se entienden** o no se ven claros en algunos paneles. Por ejemplo, la pantalla de bienvenida mostraba información general en unos gráficos que no permiten identificar claramente qué parámetros se están teniendo en cuenta. Además, todavía existían algunos campos que no estaban traducidos y aparecían en chino, cosa que provocó cierta confusión durante la evaluación. Tampoco existía uniformidad en cuanto al idioma, ya que había apartados aún en español, en vez de inglés como habíamos decidido trabajar para dar mayor visibilidad a la herramienta.

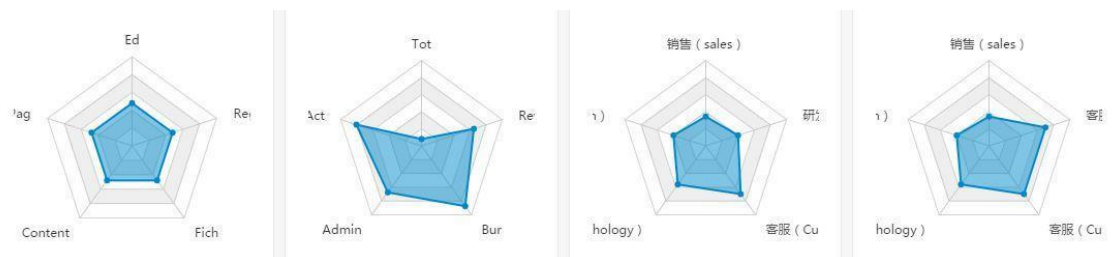


FIGURA 69. ERRORES EN EL TEXTO DE ALGUNOS GRÁFICOS

En la sección Wiki Compare se encontraba por defecto seleccionada la primera wiki disponible en nuestra lista. Ninguno de los evaluadores se dio cuenta de ese detalle, y cuando fueron a seleccionar dos wikis para compararlas vieron que, sin motivo aparente, se habían comparado 4: las dos seleccionadas por ellos, y las dos que aparecían por defecto.

En el apartado de líderes de una wiki, uno de los usuarios encontró desconcertante que no se mostrara ningún tipo de información si en ese caso concreto no existía tabla de líderes, y esperó durante unos segundos antes de darse cuenta de que la página sí había cargado pero no mostraba nada.

#### Conclusiones:

El feedback general respecto a la apariencia de la aplicación resultó positivo. Todos los usuarios encontraron fácil y atractiva la disposición de los paneles, sin elementos ni texto excesivo que quitara claridad a la interfaz. Sin embargo, había que pulir algunos

aspectos que habíamos pasado por alto durante el desarrollo de la misma, como eliminar botones sin uso, unificar el idioma de la herramienta y otros detalles que, si bien no suponían un trabajo excesivamente complejo, estaba claro que llamaban mucho la atención de los usuarios y era necesario corregir antes de presentar el proyecto final.

### 7.3.2. CALIDAD DE LA INFORMACIÓN

Dos de los tres usuarios que participaron en el proceso de evaluación accedían de manera frecuente a comunidades de Wikia, por lo que estaban familiarizados con los conceptos con los que trabajamos en la aplicación y no tuvieron demasiados problemas para entender de qué se estaba hablando en términos generales. El otro entrevistado, aunque sí habitual consumidor de otras plataformas colaborativas, no conocía las características de Wikia, por lo que su evaluación en este sentido no resultó especialmente útil. Por este motivo vamos a centrarnos en las otras dos.

#### Aspectos positivos:

A los dos usuarios les sorprendió positivamente la cantidad de información disponible. Estaban familiarizados con algunas características de Wikia, como los logros que se otorgan a los usuarios tras realizar determinadas tareas, pero ninguno de los dos sabía, por ejemplo, que los logros variaban tanto de una comunidad a otra ni de la existencia de un tablón de líderes. A parte del buscador de wikis, lo más consultado de la aplicación resultó ser el comparador, que ambos marcan como funcionalidad *top* de la herramienta. Además, otra función que les resultó útil es la que permite acceder al perfil de un usuario para ver en qué wikis participa más, consultando también la evolución a lo largo del tiempo de esas contribuciones. Esto permite hacerse una idea bastante clara del nivel de actividad de un usuario, información que no resulta fácilmente accesible desde fuera de la herramienta.

#### Aspectos negativos:

Uno de los usuarios echó en falta algún tipo de estudio sobre, por ejemplo, el tipo de wiki más consultado, entendiendo por tipo la categoría a la que pertenecen: TV, videojuegos, Libros, Películas, etc. Es cierto que esta información no la tratamos en ningún apartado, no por falta de interés por nuestra parte, sino porque la información recogida a través del parseador no contempla todos los aspectos de las wikis que nos gustaría. Lo mismo ocurre con el sexo de los miembros de la comunidad. Otra de las sugerencias estaba relacionada con la posibilidad de mantener un registro de las IPs de los usuarios con el objetivo de poder ubicar geográficamente la procedencia de los aportes. Esta opción permitiría, por ejemplo, mostrar desde dónde se hace el mayor número de contribuciones.

### Conclusiones:

Uno de los gráficos que más confusión causó fue el *Compare graph* dentro del Comparador. Este gráfico incluía en un mismo panel los datos sobre el contenido, las páginas, las ediciones y los ficheros de dos (o hasta cuatro) wikis que se estaba comparando. Mostrar una información tan distinta dentro de un mismo esquema resultó poco claro y no permitía saber muy bien cuál era el objetivo del mismo.

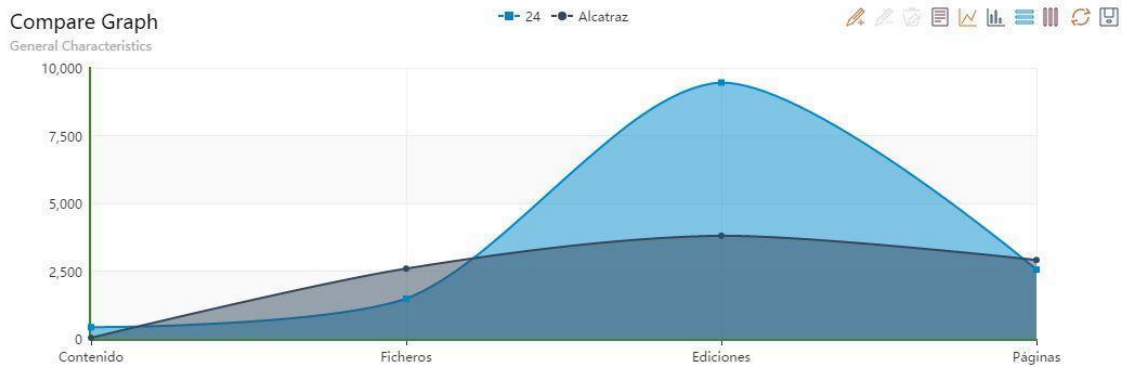


FIGURA 70. GRÁFICO QUE MUESTRA A LA VEZ CONTENIDO, FICHEROS, EDICIONES Y PÁGINAS DE VARIAS WIKIS

Esto nos llevó a la división de la información en varios sub apartados, cada uno comparando sus parámetros, y no todo mezclado en una sola figura.

En términos generales, por tanto, la evaluación resultó positiva, siendo la innovación en torno a la idea original el factor más relevante. La impresión general que nos llevamos con esta evaluación es que la información disponible era básica y no resultaba confusa, pero que necesitábamos profundizar más en el tipo de datos que queríamos mostrar. Es decir, tratar de mejorar en futuros desarrollos el tipo y calidad de información recogida por nuestro parseador.





# CAPÍTULO 8

## FUNCIONALIDAD FINAL DE CHARTSUP

En este capítulo mostraremos el resultado final obtenido tras el desarrollo de ChartsUp, con un pequeño análisis de su funcionalidad final, e incluyendo los escenarios *Key Path* descritos en el Capítulo 5.

### 8.1.1. Welcome page

Esta página proporciona enlaces rápidos a las funcionalidades principales, como el buscador y el comparador. Dicha muestra incluye el número total de ediciones, el número total de wikis incluidas, o los usuarios totales registrados.

Cada uno de los iconos enlaza con sus respectivas secciones, y además todos disponen de un botón de *info* que muestra una pequeña descripción de lo que se puede esperar de cada uno. El objetivo de esta página es permitir al usuario ver de un vistazo las principales funcionalidades de la herramienta.

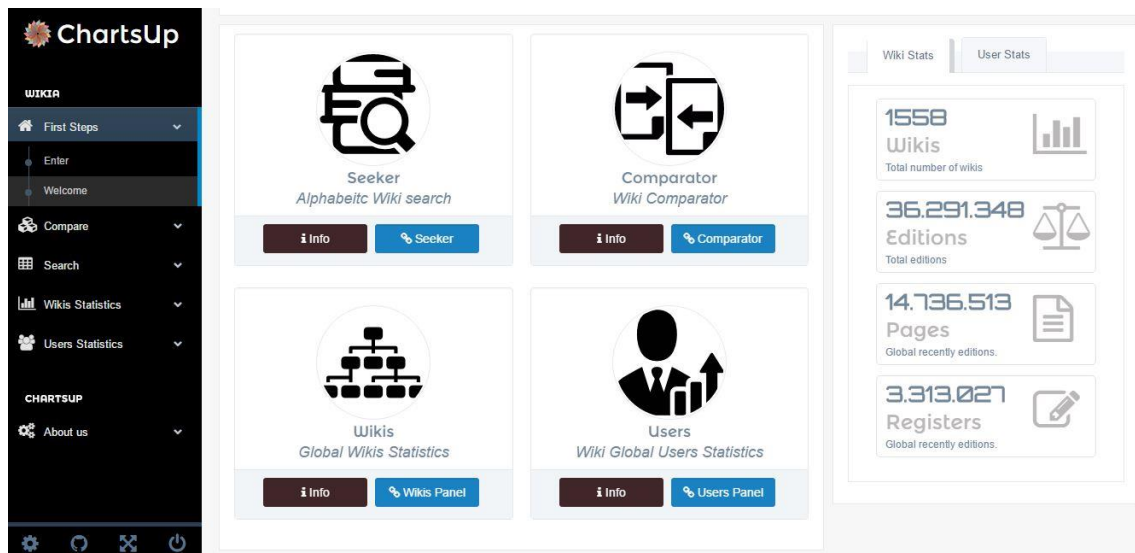


FIGURA 71. PANTALLA DE BIENVENIDA DE LA APLICACIÓN

Esta página de inicio iba a mostrar inicialmente una serie de paneles que mostraran algunos ejemplos de lo que el usuario podía encontrar navegando a través del resto de páginas de la aplicación, pero esta idea se descartó por resultar algo confuso.

### 8.1.2. Compare

El acceso al comparador muestra un formulario que permite seleccionar hasta 4 wikis mediante desplegable. Como mínimo es necesario elegir dos (las dos primeras). Más abajo encontramos las 4 wikis más comparadas hasta ese momento, en un pequeño ranking que muestra el número de veces que cada una de ellas ha sido seleccionada. Este diseño es muy parecido al que habíamos definido inicialmente en los bocetos (ver sección 5.2.4.4)

FIGURA 72. ACCESO AL COMPARADOR

Una vez seleccionadas las wikis deseadas, el botón Compare nos lleva a la página de estadísticas en la que se compara la información entre las wikis.

Esta es la funcionalidad a través de la que podemos ver reflejado el **Escenario I** planteado en el Capítulo 5: “Comparar los aspectos más importantes de dos comunidades, tales como la actividad de sus usuarios, la media de ediciones de cada wiki o las páginas por wiki con el fin de observar diferencias y tratar de mejorar alguna de ellas”.



FIGURA 73. COMPARADOR DE 3 WIKIS

En efecto, si accedemos a la página nos encontramos con una serie de gráficos que muestran información sobre, en este caso, tres comunidades distintas. Estos paneles permiten distinguir:

- Número de páginas por cada una de las wikis
- Usuarios registrados por wiki
- Ediciones y media de ediciones por wiki
- Ficheros subidos por wiki
- Cantidad de usuarios especiales (administrador, burócrata, reversionador, activo) por wiki
- Tabla de estadísticas mostrando un resumen de los anteriores paneles
- Actividad de los usuarios en función de la wiki
- Tabla de líderes para cada wiki (si existe)

Este tipo de información puede indicar que, por ejemplo, las wikis que tienen su sistema de logros habilitados tienen mucha más actividad por parte de los usuarios que las que no. Esto puede servir como orientación a un administrador de una wiki que quiera potenciar la participación en ella.

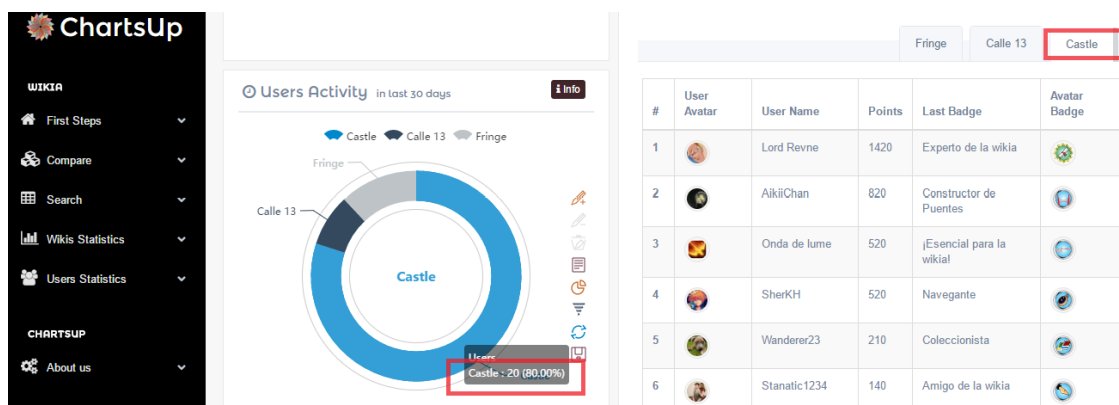


FIGURA 74. ACTIVIDAD DE USUARIOS Y TABLA DE LÍDERES DE UNA DE LAS WIKIS DEL COMPARADOR

### 8.1.3. Search

La búsqueda incluye dos opciones: general o tabla de filtrado, o alfabética. La tabla muestra por orden alfabético todas las wikis disponibles. Se divide en páginas para facilitar la navegación. La búsqueda alfabética agrupa las wikis por letra en bloques donde se indica el número total de wikis.

A través de cualquiera de las dos se puede acceder al perfil de cualquier wiki almacenada en nuestro sistema, buscando por nombre. Una vez accedemos a la wiki deseada, podemos observar una serie de información general: número de ediciones, las

páginas de esa wiki y cantidad de usuarios especiales (administradores, reversiones, burócratas y activos).

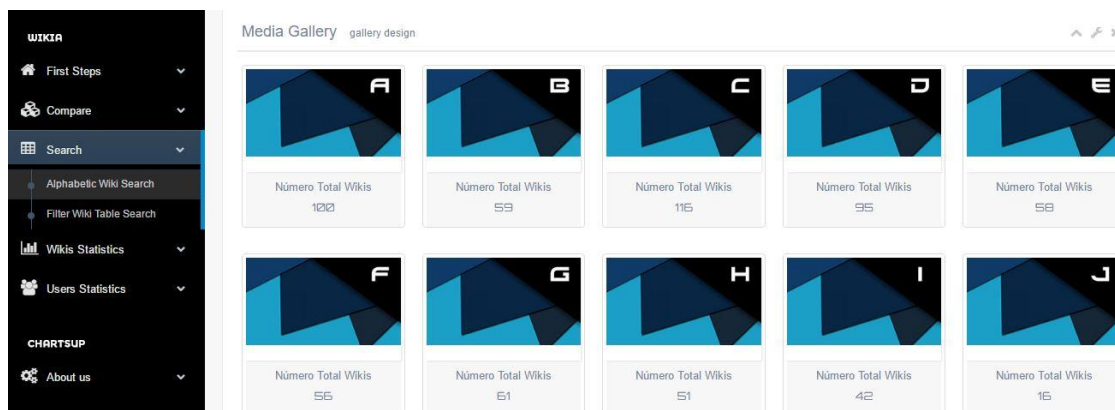


FIGURA 75. BÚSQUEDA ALFABÉTICA DE WIKIS

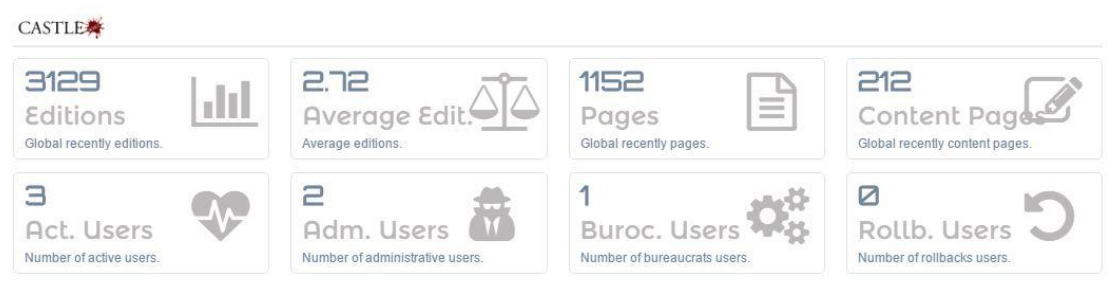


FIGURA 76. INFORMACIÓN GENERAL DEL PERFIL DE UNA WIKI

Consultando el perfil concreto de una wiki podemos “Visualizar los distintos gráficos y tablas disponibles de una wiki con el objetivo de obtener información de un determinado tema” (**Escenario III**).

#### 8.1.4. Perfil de una wiki

Lo primero que nos encontramos al acceder al perfil de una wiki es un resumen que incluye:

- Ediciones de la wiki y su media
- Número de páginas y páginas con contenido
- Cantidad de usuarios activos, administradores, burócratas y reversiones

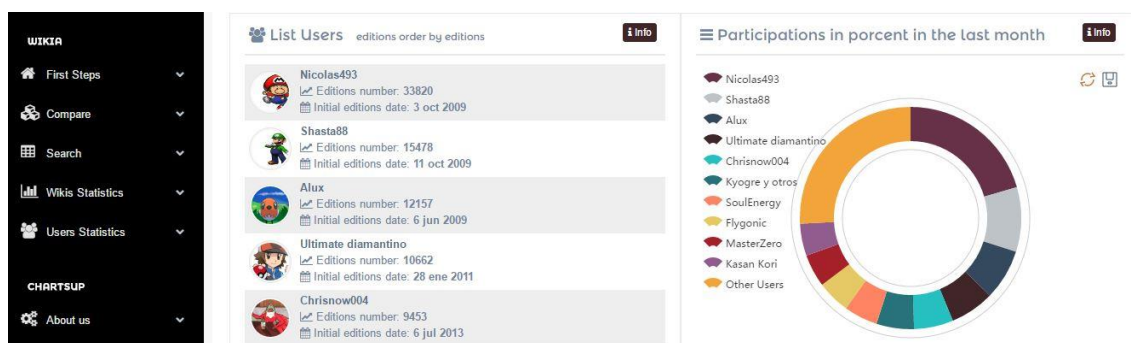


FIGURA 77. LISTA DE USUARIOS (IZDA) Y PARTICIPACIÓN (DCHA) EN LA WIKI

Más abajo aparecen dos paneles complementarios: una lista con los usuarios que más ediciones ha hecho a esa wiki, y un gráfico que permite ver en qué proporción han participado dichos usuarios.



FIGURA 78. MEDIA DE EDICIONES Y FICHEROS PARA ESA WIKI

A continuación se encuentran las cuatro gráficas que comparan distintos aspectos de la wiki (número de ediciones y media, páginas y ficheros subidos) con las 50 wikis más importantes (en relación a esas categorías) y al conjunto de todas las wikis. Esto permite ver rápidamente dónde se sitúa una wiki respecto del conjunto de las más importantes y del total.

Y por último tenemos una tabla con los últimos logros conseguidos por esos usuarios con más ediciones en la wiki que mostrábamos al principio.

### 8.1.5. Wikis Statistics

Esta funcionalidad incluye, dentro del desplegable del menú lateral, 9 categorías. Las 8 primeras hacen referencia a las páginas que disponen información sobre:

## 1. Top administrators

El primer panel dispone una lista en la que aparecen las 10 wikis con más usuarios administradores. A su derecha se encuentra un gráfico tipo anillo en el que se refleja, en el conjunto de todas las wikis, qué porcentaje de usuarios administradores es relativo a esas 10 wikis del top. Para referirnos al resto creamos la etiqueta *Other wikis*.

Gracias a la funcionalidad de la librería es posible ocultar cada una de las partes incluidas en el gráfico, por lo que, si ocultamos *Other wikis*, observamos con más claridad qué porcentaje de ese tipo de usuarios corresponde a cada wiki.

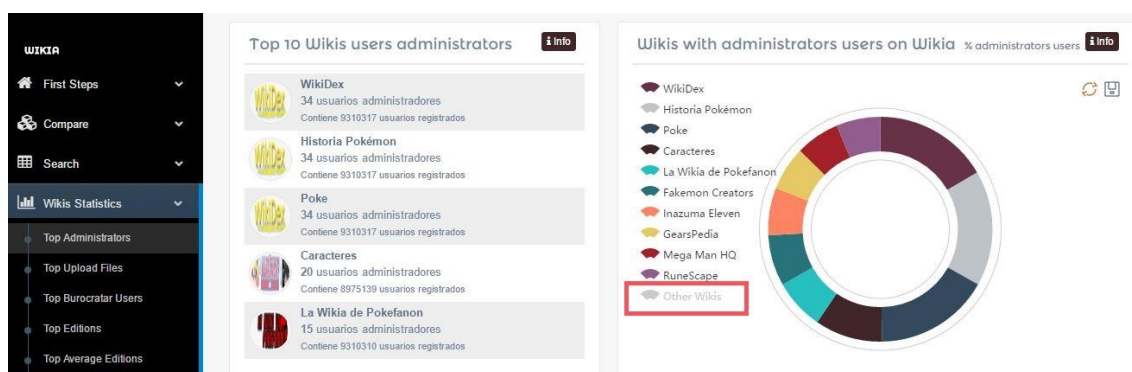


FIGURA 79. TOP ADMINISTRADORES, CON LA VISUALIZACIÓN DEL RESTO DE WIKIS DESMARCADAS

2. Top uploaded files: dispone las 10 wikis más importantes en relación con los ficheros que han sido adjuntados por usuarios (imágenes y vídeos fundamentalmente).
3. Top bureaucrats: incluye información del top 10 wikis teniendo en cuenta solo este tipo de usuario.
4. Top editions: presenta el mismo análisis que las anteriores páginas, pero la categoría es el total de ediciones para cada wiki.
5. Top average editions: muestra el top wikis que tienen una media de ediciones de contenido más alta.
6. Top rollbacks: en este caso se analizan las 10 wikis en referencia al número de usuarios de tipo reversionador.
7. Top pages: analiza las wikis en función de las páginas.
8. Top active users: muestra las 10 wikis que tienen más usuarios activos. Como ya definimos anteriormente, los usuarios activos son aquellos que han realizado algún aporte en los últimos 30 días. Esta clasificación es interna de Wikia.

Todas estas subsecciones presentan la misma disposición: una lista con el top 10 de usuarios para esa categoría, situada al lado de la gráfica tipo anillo, y una gráfica en la que se observa el top 10 wikis para la categoría correspondiente.

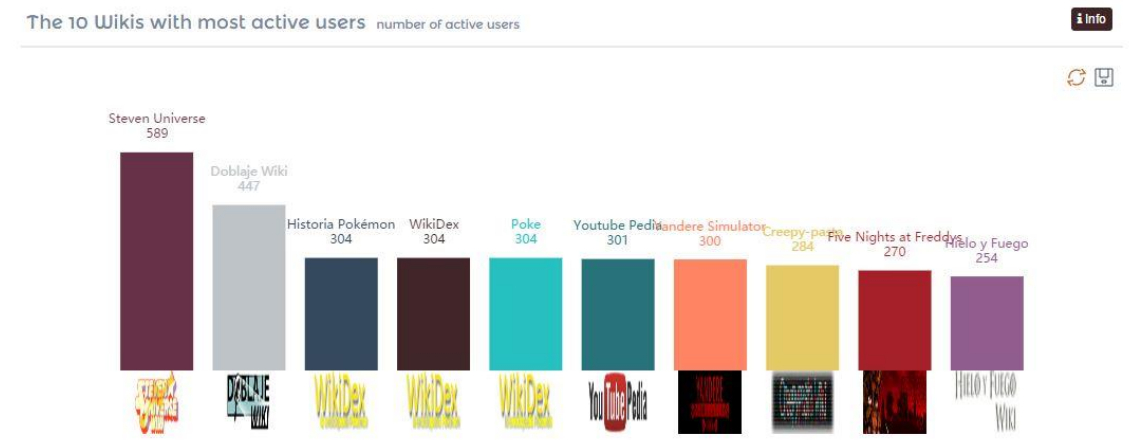


FIGURA 80. LAS 10 WIKIS CON MÁS ADMINISTRADORES

La siguiente subsección es el Customizable Scatter. Esta página permite al usuario personalizar un gráfico: mediante dos desplegables, permite seleccionar las categorías tanto para el eje X como el Y, y establecer máximos y mínimos para cada uno de los ejes.

Customizable Scatter

Axis X \*  
Number of Pages

Axis Y \*  
Number of Pages

Min Value for Axis X  
Max Value for Axis X

Min Value for Axis Y  
Max Value for Axis Y

Display!

FIGURA 81. PERSONALIZACIÓN DE LOS VALORES DEL SCATTER GRAPH

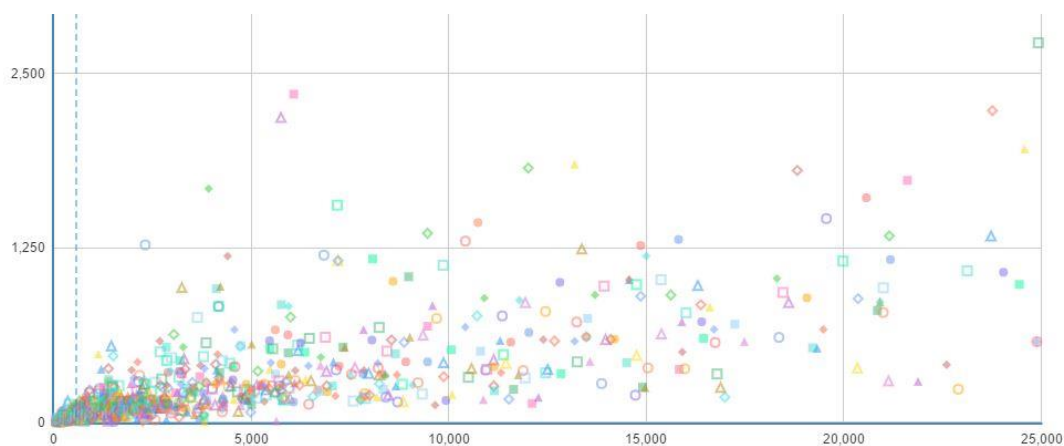


FIGURA 82. GRÁFICO DE DISPERSIÓN PARA EDICIONES (EJE X) Y PÁGINAS (EJE Y)



### 8.1.6. Users Statistics

Esta es la funcionalidad encargada de mostrar información relativa a usuarios y su actividad en el conjunto de las wikis registradas en nuestra base de datos.

El primer panel de esta sección es un gráfico que muestra la **evolución a lo largo del tiempo de las ediciones** realizadas. Esta serie temporal permite ver tanto el mínimo como el máximo año en los que se han producido esas ediciones, así como la media de ediciones.

Más abajo observamos un panel que contiene un top 10 usuarios en función de su actividad (es decir, los usuarios más activos dentro del conjunto de todas las wikis), así como una lista de los mismos.

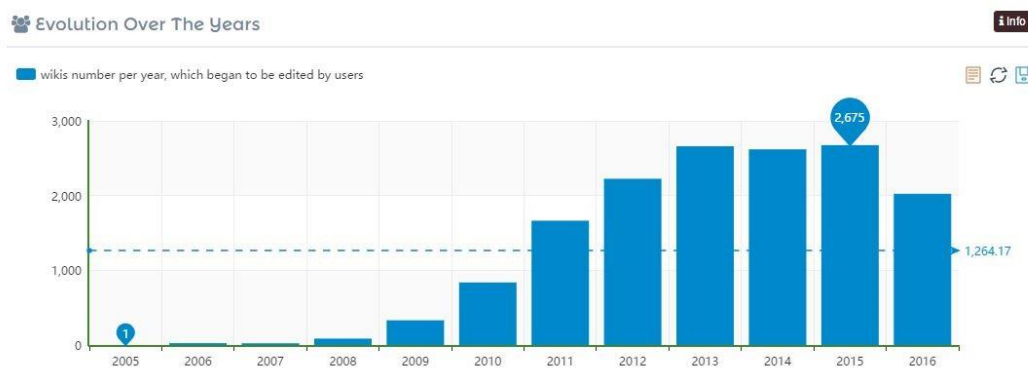


FIGURA 83. EVOLUCIÓN DE LA PARTICIPACIÓN EN WIKIA



FIGURA 84. USUARIOS CON MAYOR NÚMERO DE EDICIONES



La tabla que encontramos más abajo hace referencia a los 500 usuarios con más actividad en relación con su número de ediciones. La tabla permite realizar búsquedas por nombre de usuario, y permite acceder a todos los perfiles.

El gráfico situado a la izquierda muestra la actividad de los usuarios, agrupados por categorías (muy activo, activo, medio y bajo) dentro del conjunto de comunidades:

- Un usuario es considerado **muy activo** si tiene más de 15000 aportaciones en todas las wikis en las que participa
- Un usuario es **activo** si ha realizado entre 5000 y 15000 aportaciones
- Un usuario tiene un perfil de actividad **medio** si su número de aportaciones está entre 100 y 5000
- Un usuario tiene un perfil **bajo** de actividad si tiene menos de 100 aportaciones

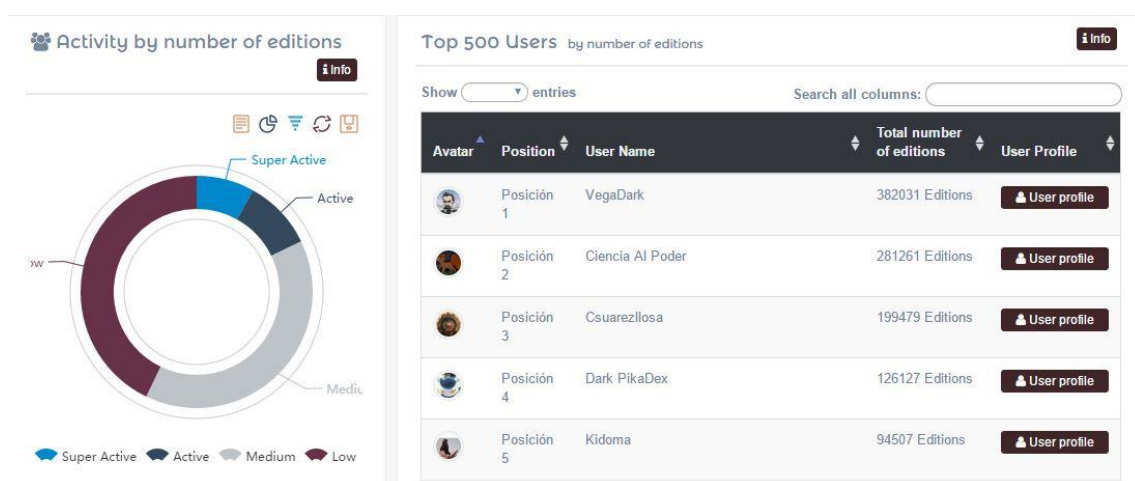


FIGURA 85. ACTIVIDAD (PROPORCIÓN) DE LOS USUARIOS (IZDA) Y LOS 500 USUARIOS MÁS ACTIVOS (DCHA),

Los siguientes dos gráficos hacen referencia a la actividad de los usuarios en función del número de wikis en las que participan. La última tabla de la página muestra así mismo los 500 usuarios más activos en referencia al número de wikis con las que se relacionan.

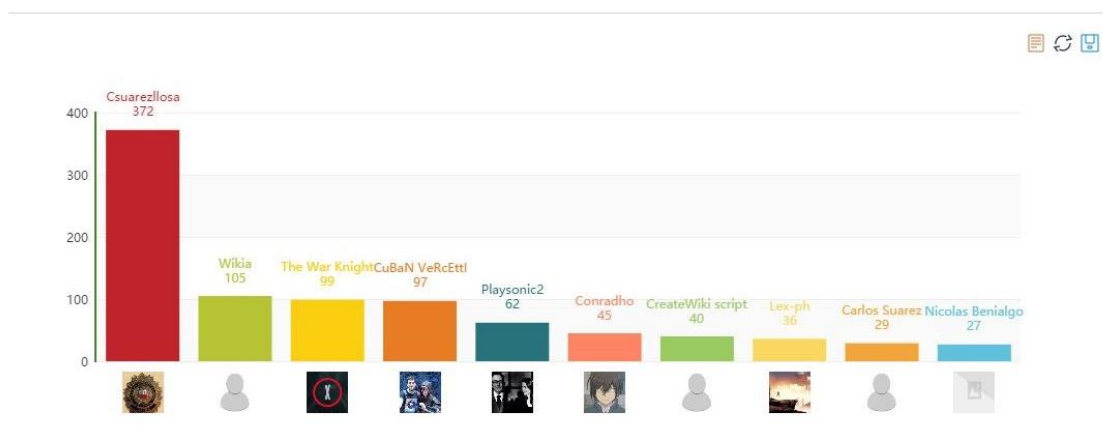


FIGURA 86. ACTIVIDAD DE LOS USUARIOS EN FUNCIÓN DEL NÚMERO DE WIKIS EN LAS QUE PARTICIPAN

Para este último caso, la actividad de los usuarios se clasifica de la siguiente forma:

- Un usuario es **muy activo** si participa en más de 20 wikis
- Un usuario es **activo** si participa en entre 5 y 20 wikis
- Un usuario tiene una actividad **media** si participa mínimo en 2 y máximo en 5 wikis
- Un usuario es **poco activo** si participa en menos de 2 wikis

El planteamiento es el mismo: permitir al usuario de la aplicación “Visualizar los distintos perfiles de usuarios más importantes para poder así analizar su comportamiento en el conjunto de varias wikis” (**Escenario II**). Tal y como se puede observar consultando estas estadísticas en la aplicación, se puede decir que sólo el 8,07% de los usuarios registrados en Wikia participan activamente. Esta conclusión es similar a las incluidas en el análisis de otras comunidades colaborativas descritas en el Capítulo 2.

### 8.1.7. Perfil de usuario



FIGURA 87. CABECERA DEL PERFIL DE UN USUARIO

Dentro de una página de usuario de Wikia encontramos como cabecera información acerca del número de ediciones totales que ha hecho (sin tener en cuenta si son o no de la misma wiki), los puntos totales que suma por todos los logros conseguidos y el número de wikis en las que participa.

También podemos consultar en esta página la evolución a lo largo de los años de las aportaciones que ha hecho el usuario hasta la actualidad, así como el porcentaje de actividad para cada una de las wikis.

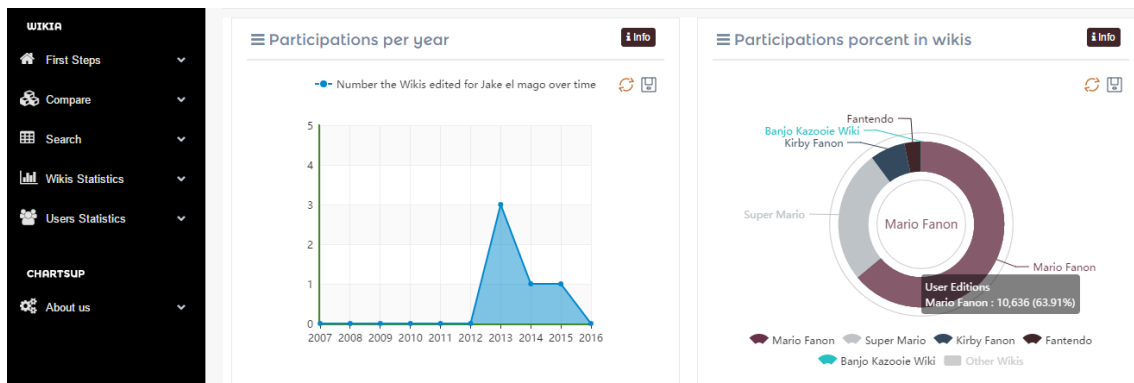


FIGURA 88. PARTICIPACIÓN DE UN USUARIO POR AÑO (IZDA) Y EN PROPORCIÓN (DCHA)

La tabla inferior se divide en dos pestañas: la primera muestra el número de ediciones por wiki, detallando la fecha en que empezó a editar, así como un enlace al perfil de esa wiki; y la segunda muestra los últimos logros conseguidos por el usuario.



FIGURA 89. TABLA CON LOS ÚLTIMOS LOGROS CONSEGUIDOS POR EL USUARIO

### 8.1.8. About us

Esta sección incluía en una primera versión acceso a páginas de tipo Contenido Legal, Contacto y Notificaciones. Pero, tras una reunión con los tutores, se decidió que resultaba más práctico proporcionar en esta categoría un enlace al repositorio GitHub con el proyecto, indicando pautas para su instalación y, en general, una descripción detallada de los objetivos del proyecto, con el fin de facilitar la contribución de otros usuarios al mismo.

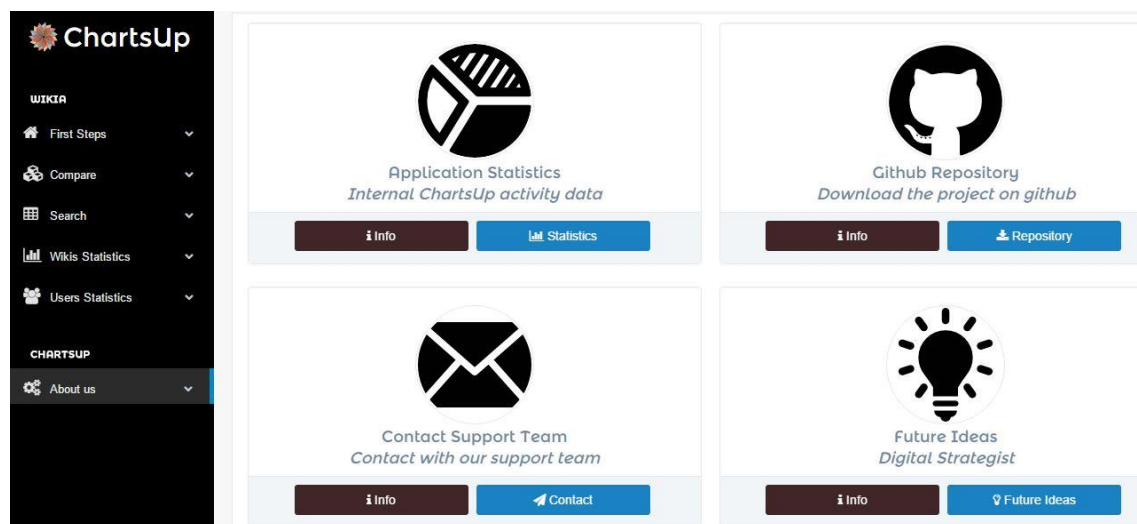


FIGURA 90. SECCIÓN ABOUT US DE CHARTSUP

Incluye también acceso al formulario de contacto (Contact Support Team) con el equipo y enlace a Future Ideas, una página en la que planteamos las futuras implementaciones de la aplicación.

La parte más importante es la de Application Statistics. Al enlazar con esta página vemos una serie de gráficas que disponen información sobre avances en el desarrollo y

contribuciones a lo largo del tiempo de la propia aplicación, dando una idea de a qué velocidad se avanza o se realizan cambios, tanto a nivel de interfaz como de base de datos.

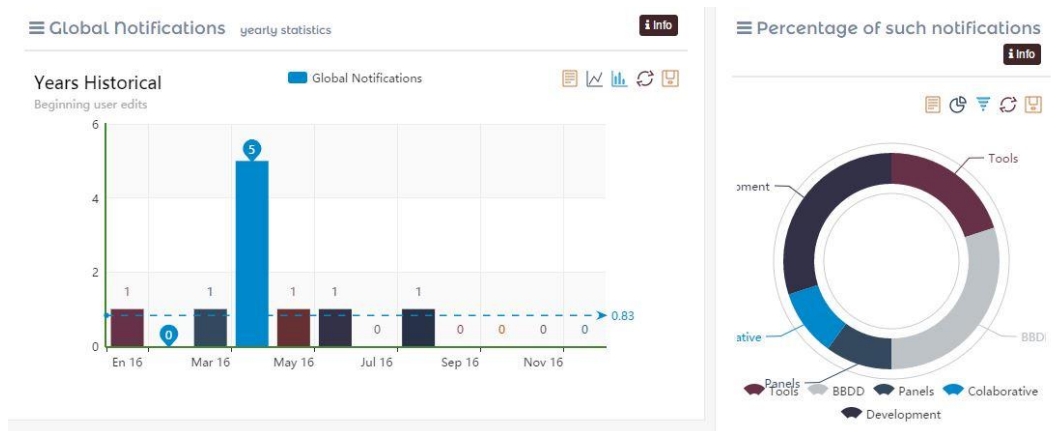


FIGURA 91. EVOLUCIÓN DEL DESARROLLO DE CHARTSUP



# CAPÍTULO 9

## RESULTADOS

Este penúltimo capítulo recoge la discusión de los resultados de todo este Trabajo de Fin de Grado, así como el reparto de tareas entre los miembros del equipo.

### 9.1. DISCUSIÓN DE LOS RESULTADOS

A lo largo de este curso académico hemos desarrollado ChartsUp, una **herramienta web cuyo objetivo es facilitar el acceso y tratamiento de información de un gran número de wikis, de forma que se pueda analizar su comportamiento de manera sencilla.**

Desde el principio se ha querido aprovechar el auge, no sólo de las herramientas de visualización, sino de la necesidad en sí misma de facilitar la comprensión de los datos que se encuentran disponibles públicamente. ChartsUp permite al usuario navegar a través de un gran número de wikis, y ver un análisis del contenido que concluimos que resultaría más útil e interesante: compararlas, analizar cada una individualmente y hacer uso de otros parámetros registrados en nuestra base de datos.

No obstante, haciendo autocritica del trabajo realizado, vemos que el tiempo dedicado al desarrollo de prototipos interactivos con Justinmind debería haber sido mayor, aunque eso retrasara la implementación de la aplicación.

Llevar a cabo la evaluación con usuarios utilizando un prototipo de alta fidelidad, en vez de la aplicación ya en desarrollo, habría simplificado mucho la posterior tarea de corrección de errores y mejora del contenido, como indican los autores de “No sólo usabilidad” [71]. Modificar algunos aspectos de la interfaz no resultó una tarea compleja (por ejemplo la unificación del idioma). Sin embargo, si hablamos del contenido de la herramienta, el proceso se complica notablemente: modificar la forma de obtener la información (o ampliarla) implica una alteración completa o, al menos, bastante profunda, de la base de datos, tipos de consultas involucradas, etc.

Se trata de un proyecto en el que se aporta una solución de software libre cuyo enfoque no se ha explotado previamente, y estamos convencidos de que permitir a otros usuarios

participar y utilizarlo libremente a través de GitHub contribuirá a la mejora general de la herramienta.

## 9.2. REPARTO DE TRABAJO

ChartsUp es un proyecto desarrollado entre tres personas. El inicio del proyecto solo necesitó de la herramienta Wrike para gestionar tareas y establecer fechas de entrega a nivel del propio equipo. Durante los primeros meses de desarrollo de código, cada uno de los miembros del grupo trabajó en local, compartiendo los cambios significativos a través de la plataforma Google Drive. El reparto de tareas durante la mayor parte del año estaba dividido y no resultó un problema gestionar de esta forma las modificaciones en la base de datos u otros ficheros. Sin embargo, a principios del mes de Mayo el volumen de código era considerable, por lo que se creó un repositorio en GitHub al que los tres miembros teníamos acceso a través del cual continuamos gestionando los cambios en el código de cualquiera de las partes incluidas en CharsUp. Esto también facilitó la revisión del trabajo por parte de los tutores del proyecto.

La contribución al proyecto disminuyó durante los meses de Enero y Diciembre debido a exámenes.

Para el seguimiento del proyecto procuramos realizar mínimo una reunión al mes con los tutores, ya fuera con uno o con ambos, aunque a lo largo del año estas reuniones no siguieron un ritmo constante. El uso del correo electrónico era la forma más habitual de comunicación, tanto para concretar reuniones como para discutir otros aspectos del proyecto.

Todas las tareas se han llevado a cabo de forma paralela. A continuación se detallan esas tareas a nivel individual.

### 9.2.1. Alejandro Pequeño Pulleiro

#### Idea de la aplicación

Alejandro comenzó en el proyecto realizando investigación sobre tipos de librerías que permitiesen mostrar unos gráficos dinámicos e interactivos. Entre ellos encontró la librería de origen chino Echarts.js que finalmente propuso al grupo por su versatilidad. Además eligió el framework que facilitaba la creación de un panel administrativo a través de estructuras predefinidas, Gentelella, basado en Bootstrap. Este sistema ahorró tiempo gracias a las plantillas disponibles.

#### Diseño y obtención de datos

Tras las primeras reuniones, y después de haber decidido centrarnos exclusivamente en mostrar información de Wikia, se encargó de investigar herramientas que permitiesen obtener información de páginas web, con el objetivo de sacarle el máximo partido a los datos disponibles en las distintas wikis. Finalmente eligió Simple HTML DOM Parser.

Después, llevó a cabo la investigación de la estructura interna de las principales páginas de Wikia que íbamos a explotar: WikiActivity, Wiki:Estadísticas y Wiki:Leaderboard, ya que estas eran las que más datos aportaban. En paralelo, diseño la estructura de la

base de datos que se utilizaría en el proyecto mediante PhpMyAdmin. Basó el desarrollo de esta estructura en el tipo de datos que se podían obtener de las páginas especiales de Wikia.

#### Implementación de la aplicación

A principios del mes de Diciembre comenzó con la implementación de la aplicación, diseñando primero un dashboard sencillo con datos ficticios para mostrar un ejemplo del posible aspecto de la interfaz futura. Este dashboard incluía ya ejemplos de paneles de la librería de gráficos elegida, con datos de ejemplo. Hacia finales de Diciembre grabó una demo visual para enseñar a los tutores el estilo de la herramienta.

Una vez desarrollada la interfaz, implementó las funcionalidades principales de ChartsUp: página del perfil de usuario y de una wiki, buscador de wikis (sus dos opciones), comparador de 4 wikis, estadísticas globales y modificación de la página de bienvenida para que se adaptara al estilo que buscábamos para nuestra herramienta. El diseño de estas funcionalidades incluía los scripts PHP que ya obtenían información real de la base de datos y la mostraban en los distintos paneles creados inicialmente. Finalmente, se encargó de subir el proyecto a GitHub y dar estructura a los directorios.

#### Redacción de la memoria

Alejandro participó en la redacción de la memoria, definiendo las tablas y relaciones establecidas en la base de datos y concretando los problemas detectados.

### 9.2.2. Claudia Gil Navarro

#### Idea de la aplicación

Durante la fase de investigación inicial, Claudia se dedicó a investigar algunas de las plataformas colaborativas que ofrecen estándares para facilitar el trabajo con database dumps, con el objetivo era poder adaptar nuestro futuro código a varias comunidades.

#### Diseño y obtención de datos

En la siguiente fase, Claudia estudió la posibilidad de obtener información de las distintas wikis consultando los database dumps disponibles en la mayor parte de las comunidades. Finalmente, el equipo se decidió por la obtención de datos mediante *crawling*. A partir de este momento, realizó el desarrollo del Diseño Guiado por Objetivos y diseñó los prototipos iniciales de la aplicación, estableciendo su interacción.

#### Implementación de la aplicación

Claudia realizó la implementación de una serie de paneles para la aplicación con información estática, con el objetivo de estudiar la disposición y uso de los gráficos en la sección de usuario de la herramienta.

También se encargó de algunas correcciones en la interfaz, así como de la explicación de los paneles informativos de cada gráfico.



#### Evaluación con usuarios

Con la aplicación semi-funcional y datos reales, planificó la dinámica de las evaluaciones con usuarios. Tras realizar las entrevistas con los 3 participantes, hizo un análisis de las conclusiones obtenidas, centrándose fundamentalmente en los aspectos más conflictivos encontrados con el fin de corregirlos.

#### Redacción de la memoria

Para la elaboración de esta memoria procedió a realizar varios borradores para su estructuración hasta obtener un primer esqueleto de la misma. Incluyendo los avances hasta Marzo, presentó un primer borrador a los tutores. Tras las correcciones pertinentes y hasta mediados de Junio, se centró en la redacción completa de la memoria para su entrega.

### 9.2.3. Alejandro del Valle Silva

#### Idea de la aplicación

Hasta mediados de Noviembre, Alejandro investigó sobre las distintas comunidades colaborativas que podíamos incluir en nuestra aplicación. Tras barajar distintas opciones, propuso centrarnos en Wikia por la cantidad de información disponible sobre múltiples temas. Al ser una comunidad de comunidades, su estructura similar permitía establecer patrones para obtener datos.

#### Diseño y obtención de datos

Ya centrados en Wikia, investigó una de las posibles formas de obtener información de las distintas comunidades que componen el conjunto de Wikia. Consultó sus distintas APIs para identificar información concreta que se podía obtener mediante llamadas e hizo algunas pruebas mediante PHP para comprobar si resultaba viable este sistema. Por problemas de latencia, esta opción se descartó.

#### Implementación de la aplicación

Una vez gestionado el sistema de obtención de datos, Alejandro se centró en el diseño de algunos de los paneles de la aplicación:

- Visualización del perfil de una wiki,
- Visualización del perfil de un usuario,
- Las estadísticas globales de usuarios y, en concreto, el desarrollo del gráfico Scatter que permite que un usuario personalice los datos que quiere consultar y establecer los límites máximos y mínimos.



# CAPÍTULO 10

## CONCLUSIONES Y TRABAJO FUTURO

En este capítulo final de hace una reflexión de las conclusiones obtenidas tras el desarrollo de ChartsUps, y se presentan algunas líneas generales de trabajo futuro a tener en cuenta para desarrollos posteriores.

### 10.1. CONCLUSIONES

El principal objetivo de este proyecto era **investigar, experimentar y aprender sobre la visualización de información**, así como analizar el comportamiento de la **plataforma colaborativa Wikia** para sacarle el máximo rendimiento a la información disponible en sus comunidades. Este objetivo se ha conseguido mediante la implementación de una herramienta web capaz de mostrar de forma dinámica e interactiva gran cantidad de información.

En un principio comenzamos analizando la **estructura que seguían las comunidades de Wikia en español**, centrándonos en las páginas concretas de cada comunidad que ofrecían más información. Finalmente, se decidió **aprovechar el patrón de las páginas con más contenido de las comunidades** para hacer *crawling* utilizando el parseador PHP Simple HTML DOM Parser.

Una vez obtenida la información, nos centramos en encontrar una librería de gráficos que se adaptara a nuestro objetivo: permitir al usuario interactuar con la aplicación. Tras estudiar algunas librerías JavaScript nos decantamos por **Echarts.js**, menos conocida pero con una gran variedad de tipos de gráficos y, sobre todo, muy dinámica.

Después nos dedicamos a estudiar las opciones de visualización de la información de forma que el usuario entendiera lo que estaba viendo. Este proceso incluyó analizar la disposición de los paneles y **estudiar qué tipo de gráfico resultaba más útil para cada conjunto de datos**.

Finalmente desarrollamos, mediante plantillas de diseño, HTML y CSS, la aplicación web que muestra, a través de una serie de scripts PHP, la información recogida en nuestra base de datos.

La realización de este proyecto ha quedado, por tanto, marcada por dos tareas fundamentales: la comprensión de la estructura interna de las comunidades de Wikia en español, con el objetivo de obtener la mayor cantidad de datos posible; y la implementación de la aplicación web, que permite al usuario interactuar con los gráficos que recogen dicha información.

Los puntos fuertes de la aplicación son los tres mencionados con anterioridad: la búsqueda de wikis, la comparación de sus principales características y el análisis de usuarios. La versión actual de la aplicación incluye datos de más de **1500 wikis, 3 millones de usuarios registrados, 14 millones de páginas editadas y alrededor de 9000 usuarios activos**.

Gracias a la aplicación que hemos desarrollado estamos intentando **contribuir a la creación de proyectos de software libre que faciliten el acceso a una información pública** que, en numerosos casos, no resulta fácil consultar. También hemos comprobado que la viabilidad de esta herramienta proporciona una gran diversidad de posibles trabajos futuros, que explicaremos en la siguiente sección.

### Viabilidad de acceso a los datos

Hubo varios intentos de conseguir acceso a la información de las comunidades de Wikia. El primer sistema con el que trabajamos consistía en hacer llamadas a la **Wikia Content API**, que contenía toda la información pública de las comunidades. Sin embargo, tuvimos que descartar esta opción por problemas de tiempos de respuesta de las llamadas: no resultaba viable que, cada vez que un usuario hiciera *clic* en la aplicación, el sistema tardara más de un minuto en responder.

La siguiente opción consistía en analizar los **database dumps** disponibles en Wikia, pero los ficheros XML resultaban muy pesados, ya que **contenían todo el histórico de la comunidad**, y no todas las comunidades disponían de ellos.

Utilizando, como método definitivo, una librería de *crawling* resolvimos el problema del tiempo de recolección de datos, ya que en un periodo de 6 horas se consiguió disponer de toda la información contenida actualmente en la base de datos. No obstante, este sistema requirió un aprendizaje mayor que el necesario para trabajar con la API o con los dumps de Wikia.

### Complejidad de la visualización de información

Junto con el problema del acceso a los datos, el otro principal desafío que nos encontramos en el desarrollo de este proyecto apareció a la hora de presentar una correcta visualización de los datos.

Durante la implementación de los primeros paneles de la aplicación antepusimos el **atractivo visual** de los gráficos a su contenido, teniendo que depurar posteriormente los paneles, incluso llegando a modificar completamente el tipo de gráfica utilizada. Como ya reflexionamos anteriormente, utilizar una gráfica que no corresponde puede **tergiversar el sentido del análisis en el peor de los casos**, o mostrar datos poco claros en el mejor.

Por tanto, **cada uno de los paneles fue estudiado en función del objetivo a representar**: no perder de vista el sentido de la representación, **por qué y para qué se realiza la representación visual** de cada conjunto concreto de datos.

## 10.2. TRABAJO FUTURO

Esta es la versión inicial de la aplicación que nos sirve para hacernos una idea de algunas de sus funcionalidades. Como ya comentamos al principio, no debemos olvidar el objetivo que planteamos al inicio de este proyecto: incluir dentro de esta herramienta no una, sino varias plataformas colaborativas. Al tratarse de plataformas con una organización interna tan distinta unas de otras, el objetivo final consistiría en **ganar en abstracción**. No sólo a nivel de código, adaptando el parseador y la estructura interna en general, sino también a nivel de interfaz: no es lo mismo comparar dos wikis, que requieren unos gráficos determinados y adaptados a sus características, que comparar dos artículos de Wikipedia.

En esta versión para Wikia hemos podido evaluar el funcionamiento general de la idea y ponerla a prueba con algunos usuarios, tanto para mejorar aspectos de la interfaz como para comprobar el potencial de desarrollo de una aplicación de estas características.

Tenemos muchas ideas sobre qué incluir en el futuro desarrollo de la herramienta. Vamos a tratar de exponer las más importantes en los puntos detallados a continuación.

1. Teniendo en cuenta la evaluación con usuarios (ver sección 7.3), creemos que **las líneas de trabajo futuro deberían enfocarse en incluir mejoras en relación a la calidad de los datos**, ya que algunos paneles resultaron confusos para algunos de ellos. Aunque, en líneas generales, la aplicación fue recibida de forma positiva, creemos que la experiencia de usuario es mejorable en este sentido.
2. Creemos que la utilidad de la aplicación crecería exponencialmente si incluyéramos más plataformas colaborativas. Poder exportar este tipo de análisis a Wikipedia, StackOverflow o Github facilitaría el acceso a este tipo de información que puede resultar muy útil pero que no es fácilmente accesible a día de hoy. Para ello uno de los primeros pasos podría ser adaptar los datasets disponibles al estándar Open Collaboration Data Factories (ver sección 2.3)
3. Así mismo, y relacionado con el punto anterior, otro aspecto a tener presente para futuros desarrollos es **ampliar la información** con la que hacemos la visualización. Uno de los problemas más importantes con los que nos hemos tropezado a la hora de llevar a cabo este proyecto es conseguir recopilar la mayor cantidad posible de información de las wikis. Si en un futuro consiguiésemos, por ejemplo, identificar la procedencia de los usuarios que participan en estas comunidades, resultaría sencillo con los gráficos con los que contamos mostrar la localización de los usuarios más activos, qué zonas colaboran más, etc.

Como alternativa al PHP Simple HTML DOM Parser encontramos algunas herramientas para *crawlear* páginas. Una de ellas es Scrapy [72], una iniciativa

de código abierto para obtener datos de páginas web. Crawljax [73] es otra herramienta de código abierto que rastrea información de páginas web de forma automática basada en Ajax.

```
import scrapy

class StackOverflowSpider(scrapy.Spider):
    name = 'stackoverflow'
    start_urls = ['http://stackoverflow.com/questions?sort=votes']

    def parse(self, response):
        for href in response.css('.question-summary h3 a::attr(href)'):
            full_url = response.urljoin(href.extract())
            yield scrapy.Request(full_url, callback=self.parse_question)

    def parse_question(self, response):
        yield {
            'title': response.css('h1 a::text').extract()[0],
            'votes': response.css('.question .vote-count-post::text').extract()[0],
            'body': response.css('.question .post-text').extract()[0],
            'tags': response.css('.question .post-tag::text').extract(),
            'link': response.url,
        }
```

FIGURA 92. CÓDIGO SCRAPY QUE SIGUE LOS ENLACES A LAS PREGUNTAS MÁS VOTADAS DE STACKOVERFLOW Y "ARAÑA" DATOS DE CADA PÁGINA

4. También sería interesante incluir los puntos de vista de perfiles como el de sociólogos y ver qué aspectos y características valoran más y a qué tipo de información podríamos dar cabida dentro de nuestro análisis que resulte especialmente compleja de conseguir.
5. Por último, aunque desde el principio explicamos que trabajar con los datos en crudo resulta engorroso y nuestro objetivo principal era “masticar” esa información, sí sería conveniente **dar la opción al usuario de que realice él mismo su propio análisis**, permitiendo exportar todos los datos disponibles en una sección a un fichero Excel. De esta manera, si el usuario quiere analizar un comportamiento particular que no hemos contemplado en nuestra herramienta, podría fácilmente descargarse la información y trabajar con ella en un formato conocido.

En cualquier caso, este proyecto está disponible bajo una licencia MIT que cualquiera con interés puede estudiar, copiar y modificar para llevar a cabo su propio desarrollo inspirado por el nuestro.

# CHAPTER 10

## CONCLUSIONS AND FUTURE WORK

In this final chapter we make a reflection on the conclusions obtained after the development of ChartsUps, and we also present some general lines of future work to be considered for further development.

### 10.1. CONCLUSIONS

The main goal of this project was to **investigate, experiment and learn about how to display information** and analyze the behavior of the **collaborative platform Wikia** to get the most out of the data available in their communities. This goal has been achieved by implementing a web tool able to show dynamic and interactive information.

At first we started analyzing **the structure of Wikia's spanish communities**, focusing on the individual pages of each community offering more information. Finally, we decided to use the **pattern** in the most important content pages to crawl data using PHP Simple HTML DOM Parser.

Once we had the information, we focused on finding a charts library that suited our purpose: to allow the user to interact with the application. After studying some JavaScript libraries, we chose **Echarts.js**, less known but with a variety of chart types and, above all, a very dynamic library.

After that, we dedicated ourselves to considering some options for the information display so that the user understands what he is seeing. This process included analyzing the disposition of the panels and **study what type of graph was more useful for each dataset**.

Finally, we developed through design templates, HTML and CSS, the web application that shows, through a series of PHP scripts, the information in our database.

The make off this project has two main tasks: understanding the internal structure of the Wikia's spanish communities, in order to get as much data as possible; and implementation of the web tool that allows users to interact with charts that collect that information.

The strengths of the application are the three mentioned above: the wikis' search, the comparison of its main features and the user analysis. The current version of the web

tool includes data from over **1500 wikis, 3 million registered users, 14 million edited pages and around 9000 active users.**

Thanks to the application that we have developed we are trying to **contribute to the creation of open software projects that facilitate access to public information** that it is not easy to consult. We have also found that the viability of this tool provides a wide variety of possible future work, that we will explain in the next section.

#### Data access viability

There were several attempts to get access to information in Wikia's communities. The first system that worked was to make calls to the **Wikia Content API**, containing all public information about communities. However, we had to discard this option because of the response call time problems: it was not feasible that every time a user clicked on the application, the system will take more than a minute to respond.

The next option was to analyze the **database dumps** available on Wikia, but the XML files were very heavy, containing all the **historical information about the community**, and not all communities had them.

We solved the problem of accessing data by using a crawling library, as definitive method, so we were able to store all the information at the database in less than 6 hours. However, this system required more work than using the Wikia Content API or the dumps.

#### Data visualization problems

Along with the problem of data access, the other major challenge appeared when presenting a correct data visualization.

During the implementation of the first version of the panels we preferred visually appealing charts to the data content, having to debug panels and even completely change the type of graph used. As we said before, **using a wrong chart can distort the meaning of analysis in the worst case**, or display unclear data at best.

Therefore, **each of the panels was studied based on the objective to represent**: do not lose sight of the meaning of representation, **why and what the visual representation of each particular set of data is performed.**

## 10.2. FUTURE WORK

This first version of the application helps us to get an idea of some of its functionalities. As we discussed at the beginning, we should not forget the goal we set at the beginning of this project: to include not one, but several collaborative platforms. Being platforms with an internal organization so different from one another, the ultimate goal would be



to **win abstraction**. Not only at the code level, adapting the parser and internal structure in general, but also at the interface level: it is not the same to compare two wikis, which require a specific and tailored to their characteristics graphs, that compare two Wikipedia articles.

In this version we have evaluated the main functionalities and test it with some users, both to improve aspects of the interface, and to test the potential of an application of this nature.

We have many ideas about what to include in the future development of the tool. We will try to present the most important in the points listed below.

1. Taking into account the evaluation with users (see section 7.3), we believe that the lines of future work should focus on include **improvements in relation to data quality**, as some panels were confusing for some of them. Although, in general, the application was received positively, we believe that the user experience can be improved that way.
2. We believe that the usefulness of the application would grow exponentially if we included more collaborative platforms. To export this type of analysis to Wikipedia, StackOverflow or Github will facilitate access to information that may be useful but is not easily accessible nowadays. To achieve this goal, one of the first steps would be to adapt the standard datasets available to Open Collaboration Data Factories (see section 2.3)
3. In addition, and related to the previous point, another aspect to keep in mind for future developments is to **expand the information that we display**. One of the most important problems that we have encountered when carrying out this project is to gather as much information as possible. If we were able to, for example, identify the source of users participating in these communities, it would be easy to show the location of the most active users, which areas collaborate more, etc.  
As an alternative to PHP Simple HTML DOM Parser we found some tools to also crawl web pages. One is Scrappy [72], an open source initiative to get data from web pages. Crawljax [73] is another open source tool base don Ajax that tracks information from web pages automatically.
4. It would also be interesting to include the points of view from sociologists and see what aspects and features they value most and what kind of information could fit in our analysis that appears particularly complex to achieve otherwise.
5. Finally, although from the beginning we explained that work with the raw data is cumbersome and our main objective was to "chew" that information, it would be appropriate to **give the user the option to make himself his own analysis**, allowing export all the data available in a section to an Excel file. Thus, if the

user wants to analyze a particular behavior we have not included in our tool, he could easily download the information and work with it in a known format.

In any case, this project is available under a MIT license that anyone with an interest can study, copy and modify to carry out their own development inspired by ours.



# BIBLIOGRAFÍA

- [1] IBM. Bringing smarter computing to big data. 2011  
Article available at [http://public.dhe.ibm.com/software/data/sw-library/data/IBM\\_Smarter\\_Computing\\_BIG\\_DATA.pdf](http://public.dhe.ibm.com/software/data/sw-library/data/IBM_Smarter_Computing_BIG_DATA.pdf)
- [2] David McCandless. *Information is beautiful*. HarperCollins, 2009
- [3] Stephen Few. *Show me the numbers: Designing Tables and Graphs to Enlighten*. Analytics Pr, 2012
- [4] Yochai Benkler. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006  
Book available at  
[http://cyber.law.harvard.edu/wealth\\_of\\_networks/Download PDFs of the book](http://cyber.law.harvard.edu/wealth_of_networks/Download_PDFs_of_the_book)
- [5] Lisa Gansky. *The Mesh: Why the Future of Business is Sharing*. Portfolio Penguin, 2010
- [6] Rachel Botsman, Roo Rogers. *What's Mine Is Yours: The Rise of Collaborative Consumption*. Harper Business, 2010
- [7] Wikipedia. The free encyclopedia that anyone can edit  
[https://en.wikipedia.org/wiki/Main\\_Page](https://en.wikipedia.org/wiki/Main_Page)
- [8] Wikimedia Foundation. Fundraising/2013-14  
Report at [https://meta.wikimedia.org/wiki/Fundraising/2013-14\\_Report](https://meta.wikimedia.org/wiki/Fundraising/2013-14_Report)
- [9] ccMixter. Collaborate with 45,000 musicians around the world.  
<http://ccmixter.org/>
- [10] Open Source Initiative. Licenses & Standards  
<https://opensource.org/licenses>
- [11] Ortega, F. (2009). *Wikipedia: A quantitative analysis* (Doctoral dissertation, PhD thesis. Universidad Rey Juan Carlos, Madrid (Cit. on p.)).  
Few, S. (2006). *Information dashboard design* (pp. 120-206). O'Reilly.
- [12] Wikimedia Tool Labs. Compilation of charts about Wikimedia projects  
<https://tools.wmflabs.org/wmcharts>
- [13] Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008, January). On the inequality of contributions to Wikipedia. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual* (pp. 304-304). IEEE.

- [14] StackOverflow Developer Survey. 2015  
<http://stackoverflow.com/research/developer-survey-2015>
- [15] Wang, S., Lo, D., & Jiang, L. (2013, March). An empirical study on developer interactions in StackOverflow. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp. 1019-1024). ACM
- [16] Wikia. Find communities you love or create your own  
<http://www.wikia.com/fandom>
- [17] Wikia Traffic and Demographic Statistics by Quantcast  
<https://www.quantcast.com/wikia.com>
- [18] Jakob Nielsen. *The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities*. 2006  
<https://www.nngroup.com/articles/participation-inequality/>
- [19] Wikia Overview by Alexa  
<http://www.alexa.com/siteinfo/wikia.com>
- [20] Open Data Foundation.  
<http://www.odaf.org/>
- [21] ¿Dónde van mis impuestos?  
<http://dondevanmisimpuestos.es/>
- [22] World Wide Web Consortium. Linked Data  
<https://www.w3.org/standards/semanticweb/data>
- [23] Open Data Formatos. Gobierno Abierto de Navarra  
<http://www.gobiernoabierto.navarra.es/es/open-data/datos/formatos>
- [24] Open Collaboration Data Factories  
<http://census.datafactories.org/>
- [25] Open Referral. A human services data interchange specification  
<https://openreferral.org/>
- [26] OpenReferral GitHub repository  
<https://github.com/codeforamerica/OpenReferral>
- [27] Digital Ecologies Research Partnership  
<http://derp.institute/>
- [28] The Four Pillars of Effective Visualizations, by Noah Iliinsky (IBM Visual Analytics)  
Video available at <https://www.youtube.com/watch?v=nrsdgvaugKg>

- [29] Quantcast  
<https://www.quantcast.com/about-us/>
- [30] Quantcast Passes \$100M Run-Rate, Lands Senior Hires As Ads Business Booms. TechCrunch, 2013.  
Article available at <http://techcrunch.com/2013/04/17/quantcast-revenue-hiring/>
- [31] Alexa. An Amazon Company  
<http://www.alexa.com/about>
- [32] Bitergia, la empresa española de analítica open source que seduce a gigantes TIC. Silicon, 2015  
Artículo en <http://www.silicon.es/bitergia-open-source-2295572>
- [33] Cauldron. Open Source Development Analytics made easy  
<https://cauldron.io/>
- [34] Google Fusion Tables  
<https://support.google.com/fusiontables/answer/2571232>
- [35] CartoDB. The Easiest Way to Map and Analyze Your Location Data  
<https://cartodb.com/attributions/>
- [36] Tableau Software  
<http://www.tableau.com/es-es/about/mission>
- [37] D3.js. Build Data Visualizations with D3  
<https://www.dashingd3js.com/why-build-with-d3js>
- [38] Google Chart Tools. Interactive charts for browsers and mobile devices  
<https://developers.google.com/chart/>
- [39] Chart.js. Simple yet flexible JavaScript charting for designers & developers  
<http://www.chartjs.org/>
- [40] Echarts GitHub repository  
<https://github.com/ecomfe/echarts>
- [41] NodeXL: Network Overview, Discovery and Exploration in Excel.  
<http://research.microsoft.com/en-us/projects/nodexl/>
- [42] Gephi. The Open Graph Viz Platform.  
<https://gephi.org/>
- [43] Open Source Initiative. The MIT License  
<https://opensource.org/licenses/MIT>
- [44] Nicholas W. Kohn, Steven M. Smith. *Collaborative fixation: Effects of others' ideas on brainstorming*. Applied Cognitive Psychology, vol. 25, 2011.

- [45] Cooper, A. (1996). Goal-directed software design. *Dr Dobb's Journal-Software Tools for the Professional Programmer*, 21(9), 16-23.
- [46] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 28-37
- [47] Eguíluz Pérez, J. (2009). Introducción a JavaScript.
- [48] PHP Manual  
<http://php.net/manual/es/intro-what-is.php>
- [49] SQL Tutorial  
<http://www.w3schools.com/sql/>
- [50] Datamatic visualizations  
<http://datamatic.co/>
- [51] jQuery API  
<http://api.jquery.com/>
- [52] Resig, J. (2009). jquery: The write less, do more, javascript library
- [53] Bootstrap  
<http://getbootstrap.com/>
- [54] MyBalsamiq  
<https://www.mybalsamiq.com/>
- [55] Justinmind Prototyper  
<http://www.justinmind.com/>
- [56] Prototyper, el producto español para hacer prototipos de apps que triunfa en EE.UU. ABC Tecnología, 2013  
Artículo disponible en <http://www.abc.es/tecnologia/moviles-aplicaciones/20131205/abci-justinmind-prototyper-triunfo-espanol-201312042057.html>
- [57] MySQL. The World's Most Popular Open Source Database  
<http://www.oracle.com/us/products/mysql/overview/index.html>
- [58] PhpMyAdmin. Bringing MySQL to the Web  
<https://www.phpmyadmin.net/docs/>
- [59] XAMPP. Apache + MariaDB + PHP + Perl  
<https://www.apachefriends.org/es/index.html>
- [60] Sublime Text  
<https://www.sublimetext.com/>
- [61] Wrike  
<https://www.wrike.com/es/>

[62] GitHub. How people build software  
<https://github.com/>

[63] Nielsen, J. (1995). 10 usability heuristics for user interface design. *Fremont: Nielsen Norman Group*.

[64] Flowing Data. Fox News bar chart gets it wrong. 2014  
Article available at <http://flowingdata.com/2014/04/04/fox-news-bar-chart-gets-it-wrong/>

[65] Huff, D., Geis, I., & Darrell Huff, I. G. (1965). *Como mentir con estadísticas* (No. 311/H88hE).

[66] Instituto Nacional de Estadística. Tipos de gráficos  
[http://www.ine.es/explica/docs/pasos\\_tipos\\_graficos.pdf](http://www.ine.es/explica/docs/pasos_tipos_graficos.pdf)

[67] API que muestra datos genéricos del conjunto de todas las páginas  
<http://www.wikia.com/api/v1>

[68] API específica para cada wiki  
[http://api.wikia.com/wiki/Wikia\\_Content\\_API](http://api.wikia.com/wiki/Wikia_Content_API)

[69] Wikia Database dumps download  
[http://community.wikia.com/wiki/Help:Database\\_download](http://community.wikia.com/wiki/Help:Database_download)

[70] PHP Simple HTML DOM Parser Manual  
<http://simplehtmldom.sourceforge.net/manual.htm>

[71] Hassan Montero, Y., & Martín Fernández, F. J. (2003). Método de test con usuarios. *No Solo Usabilidad*, (2).

[72] Scrapy. A Fast and Powerful Scraping and Web Crawling Framework  
<http://scrapy.org/>

[73] Crawljax  
<http://crawljax.com/>



